



# 63<sup>a</sup> RBras

Reunião Anual da Região Brasileira  
da Sociedade Internacional de Biometria

## Biometria e aprendizado estatístico na era da informação

📅 23 a 25 de Maio de 2018 📍 UFPR • Curitiba • PR 🌐 [rbras.org.br/rbras63](http://rbras.org.br/rbras63)

## LIVRO DE RESUMOS





# 63<sup>a</sup> RBras | Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria

Biometria e aprendizado estatístico na era da informação

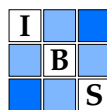
## Livro de Resumos da 63<sup>a</sup> RBras



Departamento  
de Estatística



PET  
Estatística



📅 23 a 25 de Maio de 2018  
📍 UFPR · Curitiba · PR  
🌐 [rbras.org.br/rbras63](http://rbras.org.br/rbras63)



## **Comissão Organizadora Local**

[Local Organizing Committee]

- Walmes Marques Zeviani - DEST/UFPR (Coordenador Geral)
- Wagner Hugo Bonat - DEST/UFPR (Vice-coordenador)
- Fernando de Pol Mayer - DEST/UFPR (Membro)
- Cesar Augusto Taconelli - DEST/UFPR (Membro)
- Paulo Justiniano Ribeiro Jr - DEST/UFPR (Membro)
- Silvia Shimakura - DEST/UFPR (Membro)
- José Luiz PAdilha da Silva - DEST/UFPR (Membro)
- Elias Teixeira Krainski - DEST/UFPR (Membro)
- Eduardo Vargas Ferreira - DEST/UFPR (Membro)

## **Comissão Científica**

[Scientific Commission]

- Alessandro Dal'Col Lúcio - UFSM
- Clarice Garcia Borges Demétrio - ESALQ/USP
- Fabyano Fonseca e Silva - UFV
- Guilherme J. M. Rosa - University of Wisconsin-Madison, USA
- Haavard Hue - KAUST/Arábia Saudita
- Isolde Previdelli - UEM
- Joel Mauricio Correa da Costa - The Rockefeller University, USA
- John Hinde - NUI Galway, Irlanda
- Julio da Motta Singer - USP
- Louise Ryan - UTS, Australia
- Luzia Trinca - UNESP/Botucatu
- Paulo Justiniano Ribeiro Jr - DEST/UFPR
- Renato Martins Assunção - UFMG
- Renato Ribeiro Lima - UFLA
- Ricardo Alves Olinda - UFPB
- Silvia Shimakura - DEST/UFPR
- Suely Ruiz Giolo - DEST/UFPR

## **Diretoria da RBras**

[Board of Directors of RBras]

- Presidente: Alessandro Dal'Col Lúcio (UFSM)
- Vice-Presidente: Paulo Canas Rodrigues (UFBA)
- 1ª Secretária: Taciana Villela Savian (ESALQ/USP)
- 2ª Secretário: Renato Ribeiro de Lima (UFLA)
- 1ª Tesoureira: Luzia Aparecida Trinca (IBB/UNESP)
- 2º Tesoureiro: Euclides Braga Malheiros (FCAV/UNESP)

## **Conselho Diretor da RBras**

[Council of RBras]

- Anderson Castro Soares de Oliveira (UFMT)
- Luiz Alexandre Peternelli (UFV)
- Aparecida D. P. Souza (UNESP)
- Roseli Aparecida Leandro (ESALQ/USP)
- Joel Augusto Muniz (UFLA)
- Walmes Marques Zeviani (UFPR)

## **Membros da RBras no Representative Council da IBS**

[RBras Representative Council in IBS]

- Julia Maria Pavan Soler (IME USP, SP)
- Renato Martins Assunção (UFMG, MG)
- Clarice Garcia Borges Demétrio (ESALQ/USP, SP)

## **Conselho Diretor da IBS**

[IBS Executive Board]

- President: Elizabeth Thompson (Western North American Region)
- Immediate Past President: John Hinde (British and Irish Region)

Apoio



Setor de Ciências  
Sociais Aplicadas • UFPR

Setor de Ciências  
da Saúde • UFPR

Realização



# Agropecuária

## **A acurácia da estimativa do tamanho de parcela é afetada pelo tamanho do ensaio de uniformidade**

**Dionatan Ketzer Krysczun<sup>1 3</sup>**

**Alessandro Dal'Col Lúcio<sup>2 3</sup>**

**Bruno Giacomini Sari<sup>1 3</sup>**

**Maria Inês Diel<sup>1 3</sup>**

**Tiago Olivoto<sup>1 3</sup>**

Uma das etapas do planejamento experimental é definir o tamanho de parcela que visa assegurar a precisão experimental, confiabilidade dos resultados gerados nos experimentos e redução do uso de recursos materiais, financeiros e humanos que são geralmente restritos. A estimativa do tamanho de parcela é baseada em ensaios de uniformidade, no entanto, conhece-se pouco sobre como o tamanho do ensaio de uniformidade afeta a estimativa do tamanho de parcela. Para testar esta hipótese, dois ensaios de uniformidade foram realizados com a cultura da berinjela em túnel plástico. Em ambos os tuneis foram avaliados a massa fresca de frutas e o número de frutos. Para cada ensaio, foram planejados 25 ensaios de uniformidade de diferentes tamanhos por fileira de cultivo e colheita (individual e agrupada). Para cada ensaio de uniformidade planejada, o procedimento bootstrap foi utilizado para estimar 3000 tamanhos de parcela pelo método de curvatura de coeficiente máximo de variação e calculado a média e a amplitude do intervalo de confiança de 95%, pela diferença entre os percentis 97,5% e 2,5%. O tamanho do ensaio de uniformidade influencia a estimativa do tamanho de parcela, pois ensaios com baixo número de unidades experimentais básicas apresentam alta variabilidade experimental e estimativas pouco precisa. Para estimativa do tamanho de parcela em experimentos com berinjela são necessários ensaios de uniformidade compostos por 16 unidades experimentais básicas para avaliar a variável massa fresca de frutos e 14 unidades experimentais básicas a variável número de frutos.

**Palavras-chave:** *Berinjela. Bootstrap. Planejamento experimental. Precisão experimental.*

---

<sup>1</sup> Estudante de doutorando em Agronomia, Departamento de Fitotecnia, UFSM, Santa Maria, RS. E-mail: diona1994@hotmail.com.

<sup>2</sup> Professor orientador, Departamento de Fitotecnia, UFSM, Santa Maria, RS.

<sup>3</sup> Os autores agradecem ao CNPq pelo apoio financeiro por meio das bolsas de pós-graduação e produtividade em pesquisa.



# Abordagem bayesiana no ajuste de modelos de regressão segmentada com platô de resposta: Uma aplicação aos dados de teor de Zinco (Zn) na tíbia de frangos

Rick Anderson F. Mangueira <sup>1 2</sup>

Taciana Villela Savian <sup>1</sup>

Na produção animal é importante manter níveis adequados de nutrientes nas rações, visando evitar desperdícios econômicos e principalmente o excesso nutricional na dieta animal, pois tal excesso é eliminado nas fezes e urina. Os modelos de regressão segmentada com platô de resposta são frequentemente utilizados para identificar exigências nutricionais adequadas para diversas espécies animais. A vasta utilização desses modelos deve-se ao fato de apresentarem um parâmetro que indica a dosagem ideal de um nutriente, de acordo com a variável resposta estudada. No presente trabalho foi realizado o ajuste dos modelos quadrático com platô de resposta (MQP) e não linear com platô de resposta (MNL), para identificar as exigências nutricionais de zinco (Zn) de frangos de corte. A estimação dos parâmetros dos modelos estudados foi realizada por meio da abordagem bayesiana. As aves utilizadas para o estudo foram submetidas a 9 dosagens de Zn em ppm na ração, sendo a variável resposta o teor de zinco encontrado na tíbia dos frangos. Além disso, foi utilizado para o ajuste, a média das repetições de cada dosagem realizada. Os dois modelos ajustados apresentaram um ajuste satisfatório, sendo que o modelo MQP apresentou uma dose média de Zn de 81,36 ppm para estabilização do teor de Zn na tíbia dos animais, enquanto que com o modelo MNL obteve-se um valor de Zn adequado na ração de 68,43 ppm. Além disso, o MQP apresentou um BIC de 28,45, e o MNL BIC = 15,54. Sendo assim, o MNL parece mais adequado para a amostra analisada.

**Palavras-chave:** *Estimação bayesiana, Platô de resposta, Regressão segmentada.*

---

<sup>1</sup>ESALQ-USP. e-mail: [rickanderson0310@yahoo.com.br](mailto:rickanderson0310@yahoo.com.br)

<sup>2</sup>Agradecimento à CAPES pelo apoio financeiro.

---

**Abordagens clássica, sequencial e sequencial bayesiana  
na estimação de viabilidade em sementes de soja**

**Carla Regina Guimarães Brighenti**<sup>1 3</sup>

**Bruna Cardoso Braga**<sup>2 3</sup>

Para o sucesso do plantio da soja são necessárias sementes de alta qualidade. Na avaliação da qualidade fisiológica de um lote de sementes o teste de tetrazólio é uma alternativa promissora. No processo de amostragem das sementes, deve-se priorizar a redução de tempo e custos, mantendo-se a confiabilidade. Um método para obter esta redução é a utilização da amostragem sequencial. É possível ainda, incorporar informações referentes a aspectos regionais e sazonais por meio de técnicas Bayesianas utilizando a distribuição *a priori*. Objetivou-se empregar as abordagens clássica, sequencial e sequencial bayesiana para estimar a proporção de viabilidade de germinação, através do teste de tetrazólio em lotes de sementes da soja. Utilizou-se 200 sementes de nove lotes de produtores da região do Sul de Minas para o teste clássico. O percentual médio de viabilidade de sementes de soja foi de 86,4; 87,7 e 88,3% para as abordagens clássica, sequencial e sequencial bayesiana, respectivamente. As análises sequenciais permitiram reduzir o tamanho amostral em mais de 50%, sendo utilizadas 80 sementes para abordagem sequencial e 98 para a sequencial bayesiana. Na comparação entre as três abordagens, houve concordância em 100% dos lotes avaliados pelo critério de tolerância e alta correlação entre as estimativas. A coerência entre os resultados evidencia a possibilidade do uso tanto da técnica sequencial quanto a sequencial bayesiana, utilizando menor quantidade de sementes.

**Palavras-chave:** amostragem, distribuição beta, *Glycine max* L..

---

1 DEZOO – UFSJ/ PGEE-UFLA. e-mail carlabrighenti@ufs.edu.br

2 DEZOO – UFSJ.

3 Agradecimento a FAPEMIG pelo apoio financeiro.

# Abordagens para contornar heterogeneidade de variâncias: uma aplicação em dados de crescimento em diâmetro para mudas de eucalipto

Giovana Fumes <sup>1 5</sup>

Clarice Garcia Borges Demétrio <sup>2 5</sup>

José Eduardo Corrente <sup>3 5</sup>

Juliane Fumes Bazzo <sup>4 5</sup>

Este estudo teve como objetivo propor diferentes abordagens para contornar o problema de heterogeneidade de variância presente em dados de crescimento em diâmetro de mudas de eucalipto. Os dados foram provenientes de um delineamento inteiramente casualizado, conduzido em um viveiro do Departamento de Recursos Naturais - Ciências Florestais, FCA, Botucatu, para comparar oito tratamentos diferentes, sete deles formados por diferentes proporções de lodo de esgoto e casca de arroz carbonizada e o oitavo tratamento formado pelo composto padrão utilizado no viveiro. Para cada tratamento foram realizadas quatro repetições, com vinte e quatro mudas por repetição. A variável diâmetro foi medida em quatro tempos: após 105, 120, 135 e 150 dias. Essa variável foi analisada por meio de três diferentes modelos utilizando modelos mistos e o modelo Box-Cox normal misto. Como resultado, tem-se que as abordagens propostas apresentaram bons resultados para contornar o problema de heterogeneidade de variâncias, sendo o melhor deles o modelo misto com efeito aleatório de parcela aninhado a subparcela no intercepto, considerando heterogeneidade de variâncias ao longo do tempo. Adicionalmente, o composto padrão apresentou o melhor crescimento médio/mediano quando comparado aos demais tratamentos.

**Palavras-chave:** *Box-Cox Normal, Modelo Misto, Subamostras.*

---

<sup>1</sup>Departamento de Ciências Exatas, ESALQ-USP. e-mail: [gifumesbtu@gmail.com](mailto:gifumesbtu@gmail.com)

<sup>2</sup>Departamento de Ciências Exatas, ESALQ-USP.

<sup>3</sup>Departamento de Bioestatística, IB-UNESP.

<sup>4</sup>Departamento de Ciências Florestais, FCA-UNESP.

<sup>5</sup>Agradecimento a CAPES pelo apoio financeiro.

## Ajuste do modelo de Gompertz para cultivares de alface em função da soma térmica acumulada

Milena Pacheco<sup>1</sup>

Alberto Cargnelutti Filho<sup>2</sup>

Fernanda Carini<sup>3</sup>

Jéssica Andiar Kleinpaul<sup>4</sup>

Andréia Procedi<sup>5</sup>

O objetivo deste trabalho foi ajustar o modelo de Gompertz para massa de matéria fresca de duas cultivares de alface em função da soma térmica acumulada. Foi conduzido um experimento em ambiente protegido tipo abrigo na área experimental do Departamento de Fitotecnia da Universidade Federal de Santa Maria, no outono/inverno de 2017, com as cultivares de alface do tipo americana Gloriosa® e Grandes Lagos®. Após o transplante, a cada sete dias coletaram-se seis plantas, e avaliaram-se os caracteres massa de matéria fresca de folhas comerciais (MMFFC), massa de matéria fresca de folhas não comerciais, massa de matéria fresca do caule e massa de matéria fresca das raízes. A partir destes dados calculou-se a massa de matéria fresca total (MMFT) da planta. O modelo de Gompertz foi ajustado por meio da equação  $y = a \exp[-\exp(b - cx)]$  para as variáveis dependentes (MMFFC e MMFT) em função da variável independente (soma térmica acumulada). Para cada modelo, foi calculado o ponto de inflexão. Os cálculos foram executados com o auxílio dos programas Statistica® e Microsoft Office Excel®. Os avaliadores de qualidade do ajuste foram o coeficiente de determinação ajustado ( $R^2_{aj}$ ) e o critério de informação de Akaike (AIC). O modelo de Gompertz apresentou ajuste satisfatório para a MMFFC e a MMFT para as cultivares Gloriosa® e Grandes Lagos®, ou seja, apresentou AIC próximo a zero e  $R^2_{aj}$  próximo a um.

**Palavras-chave:** *Lactuca sativa L.*, Matéria fresca, Modelagem.

<sup>1</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista FIT/BIT/UFSM.

<sup>2</sup> Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com. Bolsista de Produtividade em Pesquisa IC-CNPq - Processo: 305057/2013-8

<sup>3</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>4</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CNPq.

<sup>5</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PROBIC/FAPERGS/UFSM.

Agradecimentos: Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), pelas bolsas concedidas.

## Ajuste do modelo von Bertalanffy para caracteres produtivos de cultivares de alface

**Fernanda Carini<sup>1</sup>**

**Alberto Cargnelutti Filho<sup>2</sup>**

**Ismael Mario Marcio Neu<sup>3</sup>**

**Rafael Vieira Pezzini<sup>4</sup>**

**Milena Pacheco<sup>5</sup>**

O objetivo deste trabalho foi ajustar o modelo von Bertalanffy, para caracteres produtivos de duas cultivares de alface em função da soma térmica acumulada. Foi conduzido um experimento em ambiente protegido do tipo abrigo, com as cultivares Pira Verde® e Stella®, em Santa Maria, Rio Grande do Sul, Brasil, durante o período de outono-inverno de 2017. Para o cultivo foram utilizadas técnicas de cultivo sem solo, do tipo fechado, com calhas e vasos de 3 litros, preenchidos com substrato areia. A cada sete dias foram coletadas seis plantas e avaliada a massa de matéria fresca de folhas comerciais, em g planta<sup>-1</sup>, ou seja, aquelas que não apresentavam danos causados por pragas ou sintomas de senescência. Também foi avaliada a massa de matéria fresca total (g planta<sup>-1</sup>), constituída pelas massas frescas de folhas comerciais, não comerciais, caule e raízes. O modelo é dado pela seguinte equação  $y = a (1 - b \exp(-cx))^3$ , em que  $a$  assíntota,  $b$  razão entre os valores iniciais e o  $c$  a taxa de crescimento. Foram calculados os pontos de inflexão, de aceleração máxima e de desaceleração máxima. Os avaliadores de qualidade do ajuste foram o coeficiente de determinação ajustado e o critério de informação de Akaike. Os cálculos foram realizados com auxílio dos softwares Microsoft Office Excel® e Statistica®. O modelo de von Bertalanffy apresentou ajuste satisfatório, próximo a um para coeficiente de determinação ajustado e a zero para critério de Akaike, para os dois caracteres das cultivares Pira Verde® e Stella®.

<sup>1</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>2</sup> Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com. Bolsista de Produtividade em Pesquisa 1C-CNPq - Processo: 305057/2013-8

<sup>3</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>4</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>5</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista FIT/BIT/UFSM.

Agradecimentos: Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelas bolsas concedidas.



**Palavras-chave:** *Lactuca sativa L., Matéria fresca, Modelos de crescimento.*

## Algumas relações entre fatores influentes da análise de semivariogramas

Enio Junior Seidel<sup>1 4</sup>

Elodio Sebem<sup>2 4</sup>

Angela Pellegrin Ansuj<sup>3 4</sup>

Muitos aspectos do processo de amostragem espacial ainda necessitam ser melhor explorados e como estes aspectos impactam na dependência espacial. Neste sentido, o objetivo do trabalho é analisar as relações entre fatores de campo e algumas estimativas geoestatísticas. Foram coletados 30 artigos com um total de 295 atributos de solo, diretamente do portal de periódicos Scientific Electronic Library Online (SciElo Brasil), entre 2012 e 2016, para compor o banco de dados. Foram obtidos 30 variáveis/fatores de campo que podem influenciar nos aspectos da amostragem e da avaliação da variabilidade espacial. Primeiramente, foi realizada análise descritiva dos dados. Após, para comparar o comportamento das medianas do grau de dependência espacial (GDE) em distintos fatores de campo foram aplicados o teste de Mann-Whitney e o teste de Kruskal-Wallis, com respectivo post-hoc teste de Dunn, a 5% de probabilidade. A razão entre contribuição e patamar (%), que indica o grau de dependência espacial (GDE), mostra que a dependência espacial média foi de 69,29%, quase idêntico ao valor da mediana (69,95%), gerando, na média, uma classificação de dependência moderada. Com base nas comparações de medianas de GDE realizadas, a dependência espacial é maior nas camadas profundas de solo, em estudos com malhas irregulares de amostragem, nas maiores máximas distâncias de amostragem, com menores tamanhos de amostra e para solos do tipo Espodossolo.

**Palavras-chave:** *Geoestatística, Análise variográfica, Dependência espacial.*

---

<sup>1</sup> Departamento de Estatística – UFSM. e-mail enioseidel@gmail.com

<sup>2</sup> Colégio Politécnico – UFSM.

<sup>3</sup> Departamento de Estatística – UFSM.

<sup>4</sup> Agradecimento ao CNPq pelo apoio financeiro.

## Análise de agrupamento de semeadoras manuais

Patricia Peres Araripe <sup>1 3</sup>

Silvio Sandoval Zocchi <sup>2 3</sup>

A semeadora manual é uma ferramenta bastante utilizada por pequenos e médios agricultores em diversos países da África, Ásia e de outros na América Latina. Esse fato decorre devido a diminuição das exigências de trabalho no campo, aumento do desenvolvimento sustentável e a minimização do distúrbio do solo. Alguns trabalhos foram realizados com o intuito de comparar semeadoras manuais de acordo com a distribuição do número de sementes depositadas por golpe utilizando medidas de posição e dispersão, porém duas ou mais semeadoras com comportamentos distintos podem pertencer a um mesmo grupo considerando tais medidas na análise de agrupamento. Como metodologia alternativa, para a obtenção de grupos na análise de agrupamento foram analisadas, inicialmente, seis semeadoras manuais de diferentes fabricantes reguladas de modo que caíssem duas sementes por golpe utilizando como medida de dissimilaridade a J-divergence e os métodos aglomerativos hierárquicos (distância média, vizinho mais próximo e vizinho mais distante). As matrizes de distância e os dendrogramas são construídos para cada método abordado mostrando como as semeadoras foram classificadas em grupos nas diferentes etapas de forma hierárquica. Por fim, as conclusões foram obtidas por meio das análises realizadas e pelo conhecimento do pesquisador em relação ao experimento, sendo utilizado o software estatístico R como ferramenta para obter os resultados nas análises.

**Palavras-chave:** *Análise de agrupamento, J-Divergence, Semeadora manual.*

---

<sup>1</sup>LCE - ESALQ/USP. e-mail: [araripe\\_patricia@usp.br](mailto:araripe_patricia@usp.br)

<sup>2</sup>LCE - ESALQ/USP.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

---

**Análise de correlação canônica exploratória de atributos microbiológicos, físicos e químicos do solo em ambientes da Amazônia Meridional Brasileira**

**Vivian Aparecida Brancaglioni<sup>1</sup>**

**Marco Antonio Camillo de Carvalho<sup>2</sup>**

**Alan Rodrigo Panosso<sup>3</sup>**

**Carlos Tadeu dos Santos Dias<sup>4</sup>**

Entender as condições do solo em função do uso, em relação ao estado natural, apenas pela determinação e análise individual de seus atributos (físicos, químicos e biológicos) não é suficiente. Uma ferramenta importante é a análise de correlação canônica (ACC), técnica multivariada que busca maximizar as correlações existentes entre os conjuntos de variáveis existentes. O objetivo do trabalho foi aplicar a ACC para analisar a associação entre o conjunto de variáveis microbiológicas do solo e seus atributos físico-químicos, em diferentes usos na Amazônia Meridional, no município de Alta Floresta, MT.

O primeiro par de variáveis canônicas, com correlação canônica igual a 0.81, indicou que a variável canônica U1, apresentou maiores valores de carga canônica para o carbono da biomassa microbiana e biomassa microbiana, já a respiração basal do solo apresentou baixa carga canônica. Para a variável canônica físico-química do solo (V1) os maiores valores foram observados para saturação e teor de alumínio e teores de magnésio, silte, cálcio e argila do solo. Esses resultados indicam que solos de altos teores de alumínio, conseqüentemente, saturação por alumínio e argila foram aquelas que apresentaram os menores teores de Ca, Mg e silte. Foram observadas elevadas correlações canônicas entre os atributos microbiológicos e físico-químicos do solo, principalmente entre o carbono da biomassa microbiana, e os atributos químicos teor e saturação por alumínio no solo. Assim, a ACC evidenciou a sensibilidade do carbono da biomassa microbiana às alterações iniciais de atributos físico-químicos do solo oriundos dos diferentes usos do solo.

**Palavras-chave:** *Análise de correlação canônica; Qualidade do solo; Microbiológicos; Físicos; Químicos.*

---

<sup>1</sup> LCE – ESALQ/USP. e-mail: vivian.brancaglioni@usp.br

<sup>2</sup> Dep. Agronomia – UNEMAT: e-mail: marcocarvalho@unemat.br

<sup>3</sup> DMAT – FEIS/UNESP. e-mail: alanrp@mat.feis.unesp.br

<sup>4</sup> LCE – ESALQ/USP. e-mail: ctsdias@usp.br @usp.br

## **Análise multivariada: Estudo sobre a preferência de bebidas energéticas consumidas por atletas**

**Simone Daniela Sartorio de Medeiros<sup>1</sup>**

**Isabela Rótolí Micaroni<sup>2</sup>**

**Marta Regina Verruma-Bernardi<sup>1</sup>**

**Maressa de Oliveira Henrique<sup>3</sup>**

### **Resumo**

O perfil dos consumidores de bebidas energéticas é bastante variado, sendo utilizadas para diversos fins. Nas últimas décadas, a Estatística Multivariada tem sido utilizada como ferramenta em análise de testes afetivos. O objetivo do estudo foi aplicar, por meio do software R, as técnicas multivariadas de mapa de preferência e análise de variância multivariada (MANOVA) aos dados de aceitabilidade de bebidas energéticas. Realizou-se a análise sensorial dos atributos aparência, cor, aroma, sabor e aceitação global de 5 amostras (DTI, GWO, CKQ, LBR, XSH) oferecidas a 50 atletas. O mapa de preferência, baseado na análise de componentes principais, permite identificar a principal variação dos dados, obtendo-se a representação gráfica da diferença de aceitação entre produtos considerando a individualidade dos avaliadores. A MANOVA compara as diferenças de médias entre grupos utilizando mais de uma variável dependente simultaneamente, vantajosa em relação a análise univariada por considerar eventuais relações entre variáveis. Pelos mapas de preferência, a aceitabilidade foi definida para os atributos aceitação global e aroma, onde LBR foi a amostra preferida, e para sabor, onde a preferência individual foi voltada para a amostra CKQ. Quanto a MANOVA, para a maioria dos atributos, a bebida CKQ foi a que recebeu a maior média, e LBR a amostra de pior avaliação além de ser estatisticamente diferente das demais. Conclui-se que a aceitabilidade individual do avaliador não necessariamente é refletida pelo vetor de médias de aceitação, podendo gerar resultados discordantes. Cabe ao pesquisador decidir pelas suas necessidades qual técnica utilizar.

**Palavras-chave:** *Mapa de preferência, MANOVA, Análise sensorial.*

---

<sup>1</sup> Docente do DTAiSER – CCA/UFSCar, *Campus Araras/SP*.

<sup>2</sup> Graduanda do curso de Bacharelado em Engenharia Agrônômica – CCA/UFSCar, *Campus Araras/SP*. E-mail: [isabela.micaroni@hotmail.com](mailto:isabela.micaroni@hotmail.com). Agradecimento ao CNPq pelo apoio financeiro (PIBIC).

<sup>3</sup> Graduada em Bacharelado em Biotecnologia – CCA/UFSCar, *Campus Araras/SP*.



## **Análise multivariada e transformada wavelet aplicadas na modelagem da assinatura digital multiespectral da radiação foliar refletida em mudas de eucaliptos com bacteriose.**

**José Raimundo de Souza Passos<sup>1</sup>**

**Eniuce Menezes de Souza<sup>2</sup>**

**Edson Luiz Furtado<sup>3</sup>**

**Magali Ribeiro da Silva<sup>4</sup>**

**Tadeu Antônio Fernandes da Silva Júnior<sup>5</sup>**

A bacteriose foliar do eucalipto é uma importante doença em viveiros florestais, que se caracteriza por lesões foliares em mudas. Esta doença pode atingir mais de 3 milhões de mudas por ano, acarretando em perdas estimadas de aproximadamente de 2,4 milhões de reais por ano. Diante deste quadro, este trabalho tem por objetivo detectar a infecção por bactérias em mudas eucaliptos em viveiros, antes do aparecimento de sintomas visíveis nas folhas, permitindo assim, que o pesquisador tenha maior controle no combate da infestação. Para tal, foram aplicadas técnicas multivariadas de componentes principais e função linear discriminante de Fisher, como também a transformada wavelet discreta não decimada. As técnicas de componentes principais e da função linear discriminante de Fisher foram uteis para identificar a bacteriose em clones de Eucaliptos spp. com uma taxa de acerto de 75%. As transformadas wavelet discreta não decimada combinadas com modelos lineares generalizados sugerem que para diagnosticar a bacteriose, podemos trabalhar com as reflectâncias foliares com comprimentos de onda superiores a 1.126 nm. A banda do infravermelho em que a correlação de Spearman entre reflectância foliar e a severidade da bacteriose foliar foram significativas ocorreu entre os comprimentos de onda de 1.403,15 nm a 1529,04 nm. Foi ajustado um modelo linear generalizado de regressão utilizando a distribuição gama e função de ligação logarítmica que estima a severidade da bacteriose a partir dos comprimentos de onda (nm) na faixa de 1.194,016 a 1.644,467 nm.

**Palavras-chave:** *componentes principais, função linear discriminante, Xantomonas.*

<sup>1</sup> Departamento de Bioestatística. Instituto de Biociências. Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP. SP, Brasil.

<sup>2</sup> Departamento de Estatística da Universidade Estadual de Maringá-PR.

<sup>3</sup> Departamento de Produção Vegetal – FCA/UNESP, Botucatu-SP.

<sup>4</sup> Departamento de Ciência Florestal – FCA/UNESP. Botucatu-SP.

<sup>5</sup> Universidade do Sagrado Coração (USC), Bauru-SP

## Aplicação de modelos lineares generalizados na análise dos dados de mortalidade de *Apis mellifera*

Joatan Machado da Rosa <sup>1</sup>

Simone Silmara Werner <sup>2</sup>

Cristiano João Arioli <sup>3</sup>

Anderson Dionei Grutzmacher <sup>4</sup>

Marcos Botton <sup>5</sup>

O presente trabalho objetivou estudar a aplicação de modelos lineares generalizados na análise dos dados de mortalidade de abelhas após a aplicação tópica de inseticidas utilizados no controle de moscas-das-frutas. O experimento foi realizado utilizando o delineamento inteiramente casualizado com 10 repetições por tratamento. Cada repetição consistiu de uma caixa de acrílico com 10 abelhas. Os tratamentos foram a aplicação tópica de inseticidas com princípios ativos Malationa, Metidationa, Fosmete, Espinosade, Acetamiprido, Espinetoram, Espinosade, Alfacipermetrina. Foi realizada a aplicação de 1  $\mu\text{L}$  da formulação na região do pronoto de cada indivíduo, o grupo controle recebeu 1  $\mu\text{L}$  de água destilada. A mortalidade das abelhas foi avaliada após 1, 3, 6, 12, 24, 48 e 72 horas do início do experimento. Para análise dos dados foi empregado o modelo binomial com função de ligação probit, e como parte sistemática o delineamento inteiramente casualizado com sucessivas avaliações, tendo como tratamentos os diferentes produtos. O gráfico normal de probabilidades com envelope simulado foi utilizado para verificação do ajuste. Verificada a superdispersão incorporou-se um parâmetro de dispersão constante ( $\phi$ ), utilizando para estimação o método da máxima quase-verossimilhança. Para comparação entre os modelos usou-se a estatística qui quadrado (modelo binomial) e a estatística F (modelo com superdispersão - quasibinomial), sendo realizados testes sucessivos a partir do modelo minimal. Todas as análises foram realizadas utilizando o ambiente R, versão 3.2.1. O modelo binomial com parâmetro de dispersão ajustou-se aos dados e permitiu estimar o tempo letal de 50% dos insetos para cada princípio ativo avaliado. Em relação à classificação toxicológica os organofosforados Malationa e Metidationa foram altamente nocivos causando a morte de todas as abelhas.

**Palavras-chave:** *distribuição binomial, superdispersão, abelhas.*

---

<sup>1</sup>UDESC/ Lages. e-mail: joatanmachado@gmail.com,

<sup>2</sup>Epagri/EE Lages. e-mail: simonewerner@epagri.sc.gov.br

<sup>3</sup>Epagri/EESJ. e-mail: cristianoarioli@epagri.sc.gov.br,

<sup>4</sup>UFPEL. e-mail: adgrutzm@ufpel.edu.br.

<sup>5</sup>Embrapa Uva e vinho. e-mail: marcos.botton@embrapa.br

## **Aplicação do delineamento em Blocos Incompletos Balanceados para avaliação do diâmetro do colmo do Milho em resposta à diferentes métodos de inoculação**

**João Vitor Berner Pereira<sup>1 5</sup>**

**Gustavo Benincá Rodrigues<sup>2 5</sup>**

**Vinicius Leite<sup>3 5</sup>**

**Crysttian Arantes Paixão<sup>4 5</sup>**

**Sonia Purin da Cruz<sup>5 5</sup>**

O Delineamento em Blocos Casualizados (DBC) é utilizado principalmente em experimentos a campo. Para que o mesmo seja eficaz, o bloco deverá ser o mais homogêneo possível. Entretanto, em condições com muitos tratamentos, o mesmo não acontece, pois, a quantidade de parcelas que compõe o bloco é alta, aumentando o tamanho e a heterogeneidade deste. Esta problemática é contornada com o uso do Delineamento em Blocos Incompletos Balanceados (BIB), um delineamento em que apenas alguns tratamentos estão presentes dentro do bloco. O objetivo do estudo foi a verificação do diâmetro do colmo de milho (*Zea mays*) em R1, quando submetido a inoculações de rizobactérias promotoras de crescimento. A condução e implantação do experimento se deu no ano de 2017 em Frei Rogério/SC. O critério da Eficiência, que indica uma relação em porcentagem em comparação a um DBC, considera o número de blocos, parcelas, repetições, tratamentos e uma constante lambda. O lambda determina quantas vezes cada tratamento aparece com outro dentro das parcelas. Assim, foi escolhido o experimento com 10 tratamentos, dispostos em 6 repetições, totalizando 15 blocos, com lambda de 2, com parcelas de 6,5 x 4,0 m de dimensão, obtendo uma eficiência de 83,33%. Os dados foram submetidos à Análise Inter Bloco do Tipo III, ao nível de 5% de significância, com o teste de Wald Sequencial, verificando-se ausência de efeito dos tratamentos. Conclui-se que a inoculação não é necessária para o desenvolvimento inicial das plantas e o delineamento em blocos incompletos otimizou a disposição e condução deste experimento.

**Palavras-chave:** *Blocos Incompletos, Inoculação, Tratamentos.*

---

<sup>1</sup> João Vitor Berner Pereira – UFSC. e-mail jovitorbp@gmail.com

<sup>2</sup> Agradecimento a Total Biotecnologia.

---

**Aplicação do modelo de regressão beta no processo de incubação de ovos****Andreia da Silva Meyer<sup>1</sup>****Nelson José Peruzzi<sup>2</sup>****Gladys Dorotea Cacsire Barriga<sup>3</sup>****Amanda Liz Pacífico Manfrim Perticarrari<sup>2</sup>**

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). Esta variável tem sido analisada por diferentes metodologias estatísticas. Entretanto, como se trata de uma variável que se encontra no intervalo  $0 < y < 1$  pode-se modelar com modelo de regressão Beta (Krysicki, 1999). O presente trabalho tem como objetivo estudar a relação entre a taxa de eclosão e quatro parâmetros físicos dos ovos: peso do ovo (X1); Espessura da casca do ovo (X2); Esfericidade do ovo (X3) e a relação entre gema e albúmen (X4) com enfoque em Modelo de Regressão Beta, utilizando como função de ligação a logística. Uma análise de resíduos a partir de gráficos foi realizada para validação do modelo. Todas as análises foram realizadas no software R. Observou-se que a aplicação de modelos de regressão mais corretos para o tipo de dados no processo de modelagem, como o modelo regressão Beta, aumenta a precisão das estimativas e fornece uma análise de investigação mais consistente.

**Palavras-chave:** eclosão, modelos lineares generalizados.

**Bibliografia**

NARUSHIN, V.G.. ROMANOV M.N (b). Physical Characteristics of Chicken Eggs in Relation to Their Hatchability and Chick Weight. ASAE Annual International Meeting / CIGR XVth World Congress. Chicago, Illinois, USA July 28-July 31, 2002.

KRYSICKI, W. (1999). On some new properties of the beta distribution. Statistics & probability letters, 42(2), 131–137

---

<sup>1</sup> FEB – UNESP. e-mail andreia.meyer@feb.unesp.br

<sup>2</sup> FCAV - UNESP.

<sup>3</sup> FEB - UNESP

## Aplicação do teste de Tukey para experimentos com dependência espacial

Cristina Henriques Nogueira <sup>1</sup>

Renato Ribeiro de Lima <sup>2</sup>

Na análise de experimentos, uma vez apontada diferenças entre as médias dos tratamentos, surge a necessidade de identificar quais deles diferem entre si. Para tal avaliação, recomenda-se aplicar um teste de comparações múltiplas, destacando-se o teste de Tukey. Para a sua aplicação, é necessário que os erros sejam independentes. Porém, existem experimentos cujos efeitos espaciais não são neutralizados, surgindo, assim, a necessidade da incorporação dessa informação na composição do teste. Assim, o principal objetivo deste trabalho foi desenvolver uma proposta de adaptação para o teste de Tukey, de forma a incluir a informação da componente espacial ao mesmo. Essa informação espacial foi incorporada por meio de técnicas geoestatísticas. A adequabilidade do teste foi avaliada por meio da taxa de erro tipo I e o poder do teste, considerando a abordagem clássica e espacial. No processo de simulação foram consideradas diferentes configurações de dependência espacial. Verificou-se que o poder dos testes para a abordagem espacial foi superior, em todos os cenários analisados, ao poder do teste obtido por meio da abordagem clássica. Com relação às configurações espaciais, notou-se a influência do parâmetro de efeito pepita, sendo que as configurações com menores valores desse parâmetro, além manterem taxas de erro tipo I iguais aos níveis nominais estabelecidos, apresentaram divergências mais acentuadas entre o poder dos testes, espacial e clássico. Com isso, concluiu-se que a incorporação da informação espacial resultou em um teste mais eficiente para captar diferenças entre os tratamentos, quando esses eram provenientes de experimentos com dependência espacial.

**Palavras-Chave:** *Comparações múltiplas, Geoestatística, Adequabilidade.*

---

<sup>1</sup>DMAFE - IFSEMG. e-mail: [cristina.nogueira@ifsudestemg.edu.br](mailto:cristina.nogueira@ifsudestemg.edu.br)

<sup>2</sup>DES - UFLA. e-mail: [rrlima@des.ufla.br](mailto:rrlima@des.ufla.br)



## Estimação de Ratings no Futebol : Campeonato Brasileiro de 2017

Luciano Ribeiro Galvão <sup>1 3</sup>

Júlio Sílvio Sousa Bueno Filho <sup>2 3</sup>

Paulo Henrique Sales Guimarães <sup>2 3</sup>

Dentre os esportes mais praticados no mundo o futebol tem lugar de destaque. Uma das formas de analisar resultados de futebol é considerar os números de gols pró e contra o mandante em cada jogo como se seguissem a distribuição de Poisson. Neste trabalho utilizamos o primeiro turno do Campeonato Brasileiro de Futebol de 2017, por ser esta uma das mais importantes competições do esporte. Ajustamos um modelo linear generalizado aos resultados do primeiro turno e organizamos sistemas de predição para os resultados do segundo turno usando o modelo ajustado com parâmetros de ataque, defesa e mando de campo. As propriedades do ajuste foram boas, mas as da predição variaram muito de acordo com a forma de predição estabelecida.

**Palavras-chave:** *Distribuição de Poisson. Força de Ataque. Força de Defesa. Mando de Campo. Modelo Linear Generalizado.*

---

<sup>1</sup>DES - UFLA. e-mail: [lrgalvao@posgrad.ufla.br](mailto:lrgalvao@posgrad.ufla.br)

<sup>2</sup>DES - UFLA.

<sup>3</sup>Os autores agradecem à FAPEMIG, à CAPES e ao CNPq pelo apoio financeiro.

## Avaliando genótipos de mostarda com o modelo fator analítico Bayesiano

Carlos Pereira da Silva<sup>134</sup>

Luciano Antonio de Oliveira<sup>234</sup>

Cristian Tiago Erazo Mendes<sup>234</sup>

Joel Jorge Nuvunga<sup>234</sup>

Marcio Balestre<sup>234</sup>

O estudo da interação entre genótipos e ambientes é de suma importância no contexto de programas de melhoramento de plantas e ensaios agrônômicos de uma forma geral. Assim, a utilização de métodos eficientes para analisar dados provenientes de ensaios multiambientes (MET) tem motivado diversas pesquisas, objetivando maior confiança em relação a seleção e recomendação de cultivares superiores. O objetivo deste trabalho foi apresentar uma abordagem Bayesiana para os modelos Fatoriais Analíticos (FA) que são amplamente difundidos nos estudos para avaliar adaptabilidade e estabilidade em dados MET. Um conjunto de dados MET com 12 genótipos de mostarda avaliados em 6 ambientes, os ensaios experimentais foi em delineamentos de blocos casualizados com três repetições. Para analisar os dados MET utilizou-se modelo o fatorial analítico (FA) em uma abordagem Bayesiana. Os resultados evidenciaram flexibilidade na abordagem proposta para identificar genótipos mais produtivos, bem como separar subgrupos de genótipos e ambientes no que se refere ao efeito da interação GEI. Além disso, as representações gráficas para regiões de credibilidade bivariadas das cargas e escores fatoriais, juntamente com regiões de credibilidade bivariadas, ajudaram para um maior detalhamento de GEI, e sendo assim proporcionou mais informação para identificação e recomendação de genótipos.

**Palavras-chave:** *Bayesiana, Modelo Fatorial Analítico, Multiambiente.*

---

1 DES – UFLA. e-mail: ccpsilva81@hotmail.com

2 DES – UFLA.

3 Agradecimento a Capes pelo apoio financeiro.

4 Agradecimento a Fapemig pelo apoio financeiro.

## ***Bayesian networks* e meta-análise na predição da energia metabolizável de alimentos para frangos de corte**

Tatiane Carvalho Alvarenga <sup>1 3</sup>

Renato Ribeiro de Lima <sup>2 3</sup>

As equações de predição para a energia metabolizável de alimentos para aves são importantes na formulação adequada de rações pelos zootecnistas. O objetivo do presente estudo foi utilizar *Bayesian networks* (BN) e meta-análise para prever a energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn) de alimentos para frangos de corte. BN são modelos gráficos, que consistem na representação gráfica (gráfico) e probabilística (distribuições de probabilidade condicional e conjunta) das variáveis. Enquanto que a meta-análise consiste na combinação de vários estudos sobre o mesmo assunto. Assim, utilizou-se o algoritmo híbrido do pacote *bnlearn* do Software R para prever a EMAn a partir das covariáveis: proteína bruta, fibra bruta, extrato etéreo, matéria mineral, além da categoria de alimentos: energético (milho, subprodutos de milho e outros) e proteico (soja, subprodutos de soja e outros); e tipo de animal (pinto e galo). Os dados são referentes a experimentos do Brasil, no total de 568 experimentos. Foram obtidas doze equações de predição da EMAn e as mesmas foram implementadas em uma calculadora no sistema Android. Essa calculadora fornece os valores de energia de uma forma rápida e eficiente, o que contribui de maneira tecnológica e significativa para a elaboração de rações pelos zootecnistas.

**Palavras-chave:** *Inteligência Artificial. Machine Learning. Modelos gráficos.*

---

<sup>1</sup>DES - UFPA. e-mail: [tatianedex@posgrad.ufpa.br](mailto:tatianedex@posgrad.ufpa.br)

<sup>2</sup>DES - UFPA.

<sup>3</sup>Agradecimento à CAPES e FAPEMIG pelo apoio financeiro.

---

## ***Bootstrap* na comparação entre preços de produtos hortifrúti convencionais e orgânicos**

**Alfredo José Barreto Luiz<sup>1</sup>**

**Maria Aico Watanabe<sup>1</sup>**

**Lucimar Santiago de Abreu<sup>1</sup>**

Dados de levantamentos amostrais em campo podem apresentar uma estrutura desbalanceada que dificulta o uso de métodos inferenciais clássicos. Ao pesquisar preços de produtos hortifrúti, orgânicos ou convencionais, praticados por diversos estabelecimentos de Campinas/SP, surgiu o problema de fazer análise inferencial para comparar os preços médios entre grupos. Responder se o preço dos hortifrúti orgânicos é mais alto que o dos convencionais nos locais pesquisados é difícil porque existem causas de variação como: local; tipo de estabelecimento; data de coleta; diversidade em número e qualidade dos produtos oferecidos; que interferem no modelo sem que se possa isolar completamente os seus efeitos. Foi utilizada uma estratégia usando a técnica *Bootstrap* para enfrentar o desafio de comparar os preços médios de conjuntos tão diferentes. Doze pontos de venda, que incluíam feiras livres, mercados, supermercados e hipermercado, foram visitados entre julho e outubro de 2017. Foram levantados os preços de frutas e hortaliças, convencionais e orgânicas, conforme a disponibilidade em cada ponto de venda. Os preços foram coletados por ‘unidade’ em que eram oferecidos (maço, kg, unidade etc.). Para permitir a comparação, foram extraídas dez mil amostras de tamanho 10 (cesta de produtos), com reposição, de todos os 1317 preços de produtos convencionais e dos 331 preços de produtos orgânicos (independente da origem e podendo ser do mesmo produto em pontos diferentes). Os preços médios das cestas em cada grupo foram calculados e foi aplicado um teste t utilizando as variâncias *bootstrap* para comparação dos mesmos. A análise foi realizada no ambiente SAS.

**Palavras-chave:** *Análise inferencial, Bootstrapping, Teste t*

---

<sup>1</sup> Embrapa Meio Ambiente. e-mail: {alfredo.luiz; aico.watanabe; lucimar.abreu}@embrapa.br

## Comparação de modelos não lineares para o crescimento de frutos de amora preta cultivar Brazos

Édipo Menezes da Silva<sup>1 3</sup>

Victor Ferreira da Silva<sup>2 3</sup>

Jaqueline Gonçalves Fernandes<sup>2 3</sup>

Joel Augusto Muniz<sup>2 3</sup>

Tales Jesus Fernandes<sup>2 3</sup>

A amora preta é um fruto pequeno com sabor doce-ácido, que pertence ao gênero rubus. A Brazos é uma cultivar desta espécie, com sabor mais ácido do que doce e seus frutos possuindo aproximadamente oito gramas, sendo considerado grande para esta espécie. O ponto ótimo de colheita é extremamente importante para prolongar a vida deste fruto depois de colhido e as curvas de crescimento podem ser uma ferramenta útil para tornar essa decisão menos subjetiva. Os modelos de regressão não lineares desempenham um papel importante na descrição do fenômeno de crescimento. Os modelos logístico simples, duplo logístico e logístico generalizado são alguns modelos deste grupo. As medidas longitudinais foram feitas no diâmetro do fruto em 16 tempos, começando com 3 dias e terminando aos 41 dias após a antese. Os modelos foram ajustados aos dados pelo software estatístico R e a análise de resíduos feita pelos testes Shapiro-Wilk, Breusch-Pagan e Durbin-Watson, para normalidade, homocedasticidade e independência, respectivamente. Para identificação do melhor modelo os critérios de seleção AIC, BIC, DPR e R<sup>2</sup> ajustado foram usados. Após o ajuste dos modelos verificou-se que o logístico simples apresentou parâmetros não significativos. Já o modelo logístico generalizado violou o pressuposto de normalidade dos resíduos e o duplo logístico atendeu os pressupostos, seus parâmetros foram significativos e teve melhores indicadores de ajuste, sendo o indicado para descrever o crescimento do fruto de amora preta.

**Palavras-chave:** *Amora Preta, Análise de resíduos, Critério de ajuste.*

---

<sup>1</sup> DEST – UFLA. e-mail ediposvm01@gmail.com

<sup>2</sup> DES – UFLA.

<sup>3</sup> Agradecimento a CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## Comparação de modelos para descrever o crescimento corporal de novilhas leiteiras

Gabriel Machado Dallago<sup>1 5</sup>

Paulo César de Resende Andrade<sup>2 5</sup>

Darcilene de Maria Figueiredo<sup>3</sup>

Débora Santschi<sup>4</sup>

René Lacroix<sup>4</sup>

O objetivo do presente trabalho foi de comparar modelos não lineares assim como árvores de decisão de particionamento recursivo (CART) e redes neurais artificiais (RNA) na modelagem do crescimento populacional de novilhas leiteiras. O banco de dados analisado continha 3069 observações sobre o peso de 969 animais da raça Holandesa com idade variando entre 0 e 88 meses da província de Quebec, Canada. Foram ajustados os modelos de Brody, vonBertalanffy, Logístico, Gompertz, Richards e Michaelis-Menten, além dos algoritmos de *machine learning* CART e RNA. Os modelos foram ajustados no arquivo de treino (n = 2302). O ajuste foi avaliado tanto no arquivo de treino quanto no arquivo de teste (n = 767) com base no coeficiente de determinação, raiz do quadrado médio do resíduo, erro absoluto médio e a razão do desvio padrão. Os modelos de Brody e de Richards apresentaram o pior ajuste ao passo que o CART foi o modelo com o melhor ajuste. O primeiro e o último critério de partição empregado pelo CART foram 14 e 48 meses de idade, respectivamente. Estes, por sua vez, possuem relevância biológica podendo ser interpretados como idade na qual os animais apresentavam 50% do peso à maturidade e idade onde foi atingida a maturidade, respectivamente. Dessa forma, pode se concluir que o algoritmo CART pode ser utilizado para modelar o crescimento de novilhas, apresentando bom ajustamento e parâmetros biologicamente relevantes.

**Palavras-chave:** *Curva de crescimento, Novilhas leiteiras, Particionamento recursivo.*

---

<sup>1</sup> DZO – UFVJM. e-mail: gabrieldallago@gmail.com

<sup>2</sup> ICT – UFVJM.

<sup>3</sup> DZO – UFVJM.

<sup>4</sup> Valacta – Québec, Canadá.

<sup>5</sup> Agradecimentos à CAPES e ao ICT pelo apoio financeiro e à Valacta pelo banco de dados.

## Comparação dos modelos Polinomial e Logístico na descrição do crescimento de pimenta doce

Sérgio Alberto Jane <sup>1 3</sup>

Felipe Augusto Fernandes <sup>2 3</sup>

Joel Augusto Muniz <sup>2 3</sup>

Tales Jesus Fernandes <sup>2 3</sup>

A pimenta é uma hortaliça originária das Américas usada como alimento in natura, em forma de conserva, molho e de temperos. As cultivares da espécie *Capsicum annuum* são as mais cultivadas e de maior rendimento no mundo. O uso de modelos matemáticos no estudo das variáveis de crescimento ajuda a entender o desenvolvimento fisiológico das plantas sob diferentes condições de manejo. O objetivo deste trabalho foi comparar o ajuste dos modelos Polinomial e Logístico na descrição do crescimento de plantas da cultivar pimenta doce. Os dados analisados são referentes a altura de plantas tomadas no decorrer dos dias após o transplante. As análises estatísticas foram realizadas no software estatístico R versão 3.4.3. A comparação da qualidade de ajuste dos modelos foi feita por meio de análise gráfica, do coeficiente de determinação ajustado ( $R_a^2$ ) e desvio padrão residual ( $DPR$ ). A análise de resíduos indicou que estes foram normais, independentes e homocedásticos, atendendo portanto todas as pressuposições ao nível de 5% de probabilidade. Ambos modelos foram capazes de descrever a curva de crescimento da cultivar pimenta doce. O modelo Polinomial apresentou leve superioridade em relação ao Logístico, tanto pelo coeficiente de determinação como pelo  $DPR$ . No entanto, o modelo não linear Logístico por apresentar interpretação prática dos parâmetros deve ser o recomendado.

**Palavras-chave:** *Curva de crescimento, Modelos não lineares, Capsicum sp..*

---

<sup>1</sup>DES - UFPA. e-mail: [serztjane@gmail.com](mailto:serztjane@gmail.com)

<sup>2</sup>DES - UFPA.

<sup>3</sup>Agradecimento ao CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## Comparação e qualidade de intervalos de confiança bootstrap para dados com dependência espacial

Diogo Francisco Rossoni<sup>1</sup>

A avaliação da qualidade do ajuste é fundamental para qualquer análise estatística. No campo da estatística espacial, em especial a Geoestatística, as ferramentas para avaliação do ajuste do semivariograma são escassas. Geralmente, a discussão é embasada nas estimativas pontuais obtidas pelo ajuste de um modelo teórico ao semivariograma empírico. Em 2017 foi apresentado, pela primeira vez, o pacote *geotoolsR*, do software R. Dentre as várias funcionalidades do mesmo, destaca-se a possibilidade da obtenção de intervalos de confiança bootstrap para as estimativas do semivariograma. O pacote oferece cinco formas distintas para obtenção destes intervalos: método de nuvem, método de bloco, método de Solow, método de validação-cruzada e método de semivariograma. Posto isso, o objetivo deste trabalho é verificar a qualidade destes intervalos de confiança através da taxa de cobertura dos mesmos. Foram feitas diversas simulações, variando tamanho de amostra e modelo de correlação espacial. As taxas de cobertura obtidas foram comparadas com as obtidas pelo intervalo de confiança por aproximação de verossimilhança. As taxas de cobertura dos intervalos de confiança obtidos pelos cinco métodos propostos ficaram próximas das taxas nominais. O método de aproximação por verossimilhança não mostrou tal desempenho em alguns cenários, principalmente para dados com fraca dependência espacial. Ressalta-se que o método de blocos apresentou, em alguns cenários, taxas próximas a 100% e em outros, inferiores ao valor nominal. Isso pode ter ocorrido devido a quantidade de sub-blocos utilizados na reamostragem.

**Palavras-chave:** *Bootstrap; Geoestatística; Intervalo de Confiança; Taxa de Cobertura.*

---

<sup>1</sup> DES|PBE – UEM. e-mail [dfrossoni@uem.br](mailto:dfrossoni@uem.br)



## Correlação e análise de trilha em caracteres de feijão de porco

Alberto Cargnelutti Filho<sup>1</sup>

Marcos Toebe<sup>2</sup>

Bruna Mendonça Alves<sup>3</sup>

Cláudia Burin<sup>4</sup>

Cleiton Antonio Wartha<sup>5</sup>

O feijão de porco (*Canavalia ensiformis*) é uma espécie leguminosa utilizada para adubação verde e cobertura de solo. O estudo das relações lineares é importante para identificar caracteres para seleção indireta de plantas com maior produção de matéria fresca e seca. Assim, os objetivos deste trabalho foram avaliar as relações lineares entre caracteres de feijão de porco e identificar caracteres para a seleção indireta. Em um ensaio de uniformidade com área experimental de 10 m × 16 m (160 m<sup>2</sup>), aos 202 dias após a semeadura, foram selecionadas, aleatoriamente, 194 plantas. Em cada planta foram mensurados os caracteres altura de planta, diâmetro de caule, número de nós, número de folhas, número de vagens, massa verde de vagens, massa verde de parte aérea sem vagens, massa verde de parte aérea, massa seca de vagens, massa seca de parte aérea sem vagens e massa seca de parte aérea. Para cada caractere foi calculada a média e o coeficiente de variação. Foi investigada a relação linear entre os caracteres por meio de análises de correlação e de trilha. Na cultura de feijão de porco, o número de vagens tem relação linear positiva com as massas verde e seca de vagens. O número de folhas tem relação linear positiva com as massas verde e seca de parte aérea sem vagens e com as massas verde e seca de parte aérea. Os números de vagens e de folhas podem ser utilizados para seleção indireta.

**Palavras-chave:** *Canavalia ensiformis*, Relação linear, Seleção indireta.

<sup>1</sup> Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com. Bolsista de Produtividade em Pesquisa IC-CNPq - Processo: 305057/2013-8

<sup>2</sup> Departamento de Ciências Agrônômicas e Ambientais, Universidade Federal de Santa Maria (UFSM), Campus Frederico Westphalen, 98400-000, Frederico Westphalen, RS, Brasil.

<sup>3</sup> Eng<sup>a</sup>, Agr<sup>a</sup>, Dr<sup>a</sup>. Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil.

<sup>4</sup> Programa de Pós-Graduação em Engenharia Florestal, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil.

<sup>5</sup> Programa de Pós-Graduação em Fitotecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil. Bolsista CAPES.

Agradecimentos: Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelas bolsas concedidas.

## Crescimento de Pimenta Biquinho por modelos não lineares

Larissa de Oliveira Chicaglione <sup>1</sup>

Simone Daniela Sartorio de Medeiros <sup>2</sup>

Marina Rodrigues Maestre <sup>3</sup>

Adriano Sebastião dos Santos <sup>4</sup>

Victor Augusto Forti <sup>2</sup>

A pimenta biquinho *Capsicum chinense* tem sido procurada cada vez mais pelo consumidor que a escolhe por sua alta atratividade em parâmetros de sabor e baixa picância. O gênero *Capsicum* apresenta ampla variação em cor, tamanho e formato de frutos e poucas cultivares uniformes estão disponíveis. Os modelos não lineares podem ser usados em estudos para aperfeiçoar características de interesse, já que são biologicamente interpretáveis e necessitam de menos parâmetros. O objetivo do estudo foi investigar o crescimento dos frutos da pimenta biquinho em relação ao seu comprimento e largura, utilizando modelos de regressão não linear. Os dados são de um experimento conduzido em campo na Universidade Federal de São Carlos, Campus Araras - Araras/SP, entre março e novembro de 2017. Utilizou-se a variedade 37701 - Linhagem F7 (B) e, para a avaliação dos frutos, utilizou-se paquímetro digital eletrônico, sendo as medições realizadas duas vezes na semana a partir da antese, totalizando treze dias de avaliação. Os modelos não lineares estudados foram Gompertz e Logístico, com parâmetros estimados através de quadrados mínimos iterativos empregando-se o método de Gauss-Newton. Concluiu-se que, para as duas variáveis estudadas, o melhor ajuste se deu pelo modelo Logístico, que apresentou menor AIC. As taxas de crescimento estimadas pelo modelo logístico indicaram que as pimentas crescem mais rapidamente em comprimento (0,18 mm/dia), do que em relação a largura (0,15 mm/dia). As estimativas médias para o comprimento máximo dessa variedade melhorada é de 27,5 mm e largura média de 21,6 mm, sendo maiores que a variedade padrão.

**Palavras-chave:** *Comprimento, Gompertz, Largura, Logístico.*

---

<sup>1</sup>Discente em Engenharia Agrônômica - UFSCar. e-mail: [l.chicaglione@gmail.com](mailto:l.chicaglione@gmail.com)

<sup>2</sup>Docente do DTAiSeR-Ar - UFSCar.

<sup>3</sup>Docente da UEMS - Dourados/MS. - Agradecimento à Capes pelo apoio financeiro.

<sup>4</sup>Mestre em Agroecologia e Desenvolvimento Rural - UFSCar.

## Curvas de crescimento de suínos com erros autocorrelacionados de primeira ordem

Felipe Augusto Fernandes <sup>1 4</sup>

Adriele Aparecida Pereira <sup>2 4</sup>

Édipo Menezes da Silva <sup>3 4</sup>

Victor Ferreira da Silva <sup>3 4</sup>

Tales Jesus Fernandes <sup>3 4</sup>

O aumento da população mundial aponta para o crescimento do consumo de alimentos. De acordo com Guimarães et al. (2017), o consumo de carne suína tem crescido sendo a segunda proteína animal mais consumida do mundo. A criação de suínos apresenta boas perspectivas para o Brasil no mercado internacional. Conforme Fernandes et al. (2010), modelar o crescimento de determinada espécie pode facilitar a tomada de decisões relativas ao seu manejo, e para isso os modelos de regressão não linear tem sido frequentemente utilizados. O objetivo deste trabalho foi comparar o ajuste dos modelos não lineares Logístico, Gompertz e Von Bertalanffy, na descrição do crescimento do peso, em quilos, de suínos. Os dados analisados foram retirados de Oliveira et al. (2007) e são referentes ao peso de suínos fêmea, considerando a idade, em dias. As análises estatísticas foram desenvolvidas no software estatístico R. A qualidade de ajuste fornecidas pelos modelos foi comparada utilizando os seguintes avaliadores: o coeficiente de determinação ajustado ( $R_a^2$ ) e o desvio padrão residual (DPR). Nos três modelos, os parâmetros foram significativos, segundo o teste t, ao nível de 5%. Ao realizar os testes Shapiro-Wilk, Durbin-Watson e Breusch-Pagan, verificou-se que os resíduos são normais, homocedásticos e possuem autocorrelação de primeira ordem em todos os modelos, ao nível de 5% de significância. O modelo que apresentou melhor qualidade de ajuste foi o de Gompertz, com estrutura de erros autoregressivos de primeira ordem, pois apresentou valor menor do DPR, além de um maior coeficiente de determinação ajustado.

**Palavras-chave:** *Curvas de crescimento, Erros correlacionados, Suínos.*

---

<sup>1</sup>DES - UFPA. e-mail: [fernandesfelipepest@gmail.com.br](mailto:fernandesfelipepest@gmail.com.br)

<sup>2</sup>ICSA - Unifal, Campus Varginha.

<sup>3</sup>DES - UFPA.

<sup>4</sup>Agradecimento ao CNPq, CAPES e FAPEMIG pelo apoio financeiro.

---

**Decomposição em valores singulares e ajuste log-linear  
a dados agrários em tabelas de múltiplas entradas**

**Júlia Mara Campos de Souza<sup>1</sup>**

**Mariana Resende<sup>2</sup>**

**Alexandre de Oliveira Teixeira<sup>3</sup>**

**Carla Regina Guimarães Brighenti<sup>4 5</sup>**

A prolificidade num sistema produtivo de suínos é influenciada por diversos fatores incluindo o ciclo reprodutivo da matriz suína e o tamanho da leitegada. O que nos leva a poder classificar o número de partos formando tabelas de contingência em relação ao ciclo reprodutivo e o tamanho da leitegada. Para determinar a associação entre as variáveis utiliza-se técnicas multivariadas baseadas na decomposição de valores singulares (DVS), que pode ser visualmente analisado através dos biplots. Enriquecendo a análise de tabelas de contingência, um modelo log-linear é ajustado estabelecendo as relações entre as variáveis. Assim, objetivou-se, utilizando a DVS para tabelas combinadas, analisar a relação de número de partos com o tamanho da leitegada e o ciclo reprodutivo, ajustando esta a um modelo log-linear, considerando diferentes plantéis. A coleta de dados foi realizada nos plantéis de 2014 e 2015 em uma fazenda em Lagoa Dourada-MG. Obtiveram-se as coordenadas para construção dos biplots considerando a soma e a diferença entre os plantéis. Para o ajuste do modelo log-linear foi utilizada a função de ligação logit em que os valores preditos eram obtidos pela função glm do software R. Para o estudo da soma e da diferença dos planteis, os dois primeiros valores singulares fornecem mais de 80% de explicação da variabilidade. Os biplots evidenciaram o terceiro e quarto ciclo como os mais produtivos. Na diferença entre os plantéis não foi observada nenhuma associação, demonstrando padronização do manejo. Quanto ao modelo log-linear ajustado, houve alta correlação ( $r = 0,963$ ) entre os valores observados e ajustados.

**Palavras-chave:** *biplots, glm, logit*

---

<sup>1</sup> Graduanda em Zootecnia-UFSJ-PIBIC

<sup>2</sup> PGEE-UFLA

<sup>3</sup> DEZOO-UFSJ

<sup>4</sup> DEZOO-UFSJ/PGEE-UFLA. e-mail: [carlabrighenti@ufs.br](mailto:carlabrighenti@ufs.br)

<sup>5</sup> Agradecimento a FAPEMIG e ao CNPq pelo apoio financeiro

## Delineamento ótimo para curvas de crescimento de frangos de corte

Luciana Helena Kowalski <sup>1 3</sup>

Walmes Marques Zeviani <sup>2 3</sup>

A utilização de modelos estatísticos para descrever e ajustar curvas de crescimento animal assim como a aplicação de modelos ótimos vem sendo amplamente utilizados na produção animal. Sendo assim, o objetivo do trabalho foi ajustar o modelo de crescimento logístico por unidade experimental e obter o delineamento ótimo baseado no método proposto por Santos e Santos (2008). Foram utilizados dados de peso corporal de frangos de corte machos da linhagem Cobb<sup>®</sup>. Os animais foram blocados de acordo com o peso corporal inicial, eliminando-se este possível efeito nos dados. As pesagens, para verificação de peso corporal das aves, eram realizadas semanalmente. O ajuste pelo modelo logístico através da parametrização centro-escala, onde a assíntota superior do modelo, o ponto de inflexão da curva e o parâmetro de escala proporcional a taxa no ponto de inflexão, foram estimados. Foi aplicado o algoritmo sequencial para obtenção de delineamento ótimo, onde a pesquisa a priori preconizava pesagens semanais para descrição de curva de crescimento. No entanto, o delineamento ótimo obtido indica as pesagens com os animais nas seguintes idades: 1, 11, 22, 27, 32, 37 e 42 dias. Sendo assim, com a aplicação do modelo ótimo obteve-se um ganho médio em eficiência de 13% considerando o determinante de  $X^T X$ . Portanto, recomenda-se a utilização do delineamento ótimo em experimentos de avaliação de peso corporal de frangos de corte.

**Palavras-chave:** *aves, modelo logístico, peso corporal.*

---

<sup>1</sup>Graduanda DEST - UFPR. e-mail: *lucianahelenak@gmail.com*

<sup>2</sup>Professor DEST - UFPR.

<sup>3</sup>Agradecimento a Fundação Araucária pelo apoio financeiro.

## Descrição da cinética de secagem do fruto de jamelão

Edilson Marcelino Silva<sup>1 3</sup>

Thais Destefani Ribeiro<sup>2 3</sup>

Joel Augusto Muniz<sup>2 3</sup>

O fruto do jamelão, também conhecido como jambolão, jambo e azeitona do nordeste, é uma planta nativa dos trópicos e no Brasil é encontrada nas regiões Nordeste, Sudeste e Norte. O jambolão é um fruto exótico, rico em compostos antioxidantes, características nutricionais e terapêuticas. No entanto apresenta alta perecibilidade, dificultando sua produção e comercialização, necessitando de cuidados no pós colheita. Assim as técnicas de secagem constituem uma saída importante para que o fruto possa ser consumido durante todo o ano, além de auxiliar no reaproveitamento de partes muitas vezes descartadas. Este trabalho teve como objetivo estudar a cinética de secagem do extrato da fruta do jamelão nas temperaturas de 50 e 60°C, utilizando modelos de regressão não linear de Henderson e Lewis, e modelos simples e quadrático. Os dados foram obtidos experimentalmente por Silva e Souza (2017), com um total de 12 medições até o tempo de 470 minutos. A análise dos dados foi realizada utilizando software livre R. De modo geral, os modelos lineares violaram ao menos uma das suposições de normalidade e homocedasticidade dos resíduos. Os modelos de Henderson e Lewis apresentaram bom ajuste aos dados com Desvio Padrão Residual na ordem de  $10^{-2}$  e coeficiente de determinação ajustado superior a 0,98. Em ambas as temperaturas o modelo de Henderson foi considerado como mais adequado para descrever os dados pelo critério de informação de Akaike, com taxa de secagem de 0,013 g de água/min (50°C) e 0,017 g de água/min (60°C).

**Palavras Chave:** *Conservação de Alimentos; Modelos de regressão Linear; Modelos de regressão não linear.*

---

<sup>1</sup> DES – UFLA. e-mail edilsonmg3@hotmail.com

<sup>2</sup> DES – UFLA.

<sup>3</sup> Agradecimento à CAPES e à Fapemig pelo apoio financeiro.

## Descrição do acúmulo de matéria seca no feijoeiro cv. Jalo: uma comparação dos modelos não lineares Logístico e Verhulst

Laís Mesquita Silva <sup>1 3</sup>

Kelly Pereira de Lima <sup>2 3</sup>

Augusto Ramalho de Moraes <sup>3 3</sup>

Neiva Maria Batista Vieira <sup>4 3</sup>

Dâmiany Pádua Oliveira <sup>5 3</sup>

O estudo sobre o crescimento de plantas através do acúmulo de matéria seca ao longo do tempo é vantajoso, pois tal estudo se torna uma ferramenta informativa para o manejo adequado e detecção de fatores que afetam no desenvolvimento da planta, um desses fatores pode ser a densidade populacional. Desta forma, este trabalho teve como objetivo descrever o acúmulo biomassa seca do feijoeiro cv. Jalo em cinco densidades de semeadura por meio dos modelos não lineares Logístico e Verhulst. Os dados foram obtidos de um delineamento em blocos casualizados, com três repetições, no qual os tratamentos estão dispostos em esquema fatorial 2x5x7, envolvendo sistema de plantio convencional e direto, cinco densidade de semeaduras (75, 145, 215, 285, 355 mil plantas por  $ha^{-1}$ ) e sete épocas de avaliação para cada sistema de plantio, 13 a 73 e 10 a 70 dias, respectivamente. Foram analisados o acúmulo de biomassa seca no peso Total (Hastes + Folhas + Vagens + Grãos). Os modelos Logístico e Verhulst mostraram-se adequados para descrever o acúmulo de matéria seca, além disso, os pressupostos (normalidade, homogeneidade e independência) foram atendidos e pela medida de curvatura o modelo logístico foi selecionado. A partir dos modelos não lineares mencionados, foi possível estimar o peso assintótico máximo de matéria seca para cada densidade de semeadura em cada tipo de plantio, na época de ocorrência de maior acúmulo e diferenciação dos tratamentos. Assim, para a densidade de 285 mil plantas por  $ha^{-1}$  pesos máximo foram  $3401kg.ha^{-1}$  (convencional) e  $3228kg.ha^{-1}$  (direto).

**Palavras-chave:** *Phaseolus Vulgaris L.*, medidas de curvatura, biomassa seca.

<sup>1</sup>DES - UFPA. e-mail: laismesquita@hotmail.com

<sup>2</sup>DES - UFPA.

<sup>3</sup>Agradecimento ao CNPq, CAPES e FAPEMIG pelo apoio financeiro.

## **Eficiência técnica agropecuária dos municípios de São Paulo usando regressão quantílica**

**Gabriela França Oliveira<sup>1 4</sup>**

**Marcia Elaine Guimarães Lana<sup>2 4</sup>**

**Ana Carolina Campana Nascimento<sup>3 4</sup>**

**Moysés Nascimento<sup>3 4</sup>**

**Camila Ferreira Azevedo<sup>3 4</sup>**

A agropecuária é uma modalidade de produção rural baseada no cultivo de plantas e na criação de animais, constituindo-se a principal atividade econômica de diversos estados brasileiros. Diante do destaque no cenário nacional, vários estudos têm sido realizados visando analisar a eficiência deste setor. Geralmente, esses estudos baseiam-se em técnicas como Análise Envoltória de Dados e Fronteira Estocástica. Contudo, estas metodologias podem acarretar em estimativas inapropriadas da eficiência, devido à presença de heterogeneidade, o que pode ser contornado com o uso da regressão quantílica (RQ). Tendo em vista que São Paulo destaca-se como um dos estados mais importantes para a agropecuária brasileira, objetivou-se determinar a eficiência técnica da agropecuária paulista e os fatores que a influenciam, com base em modelos de RQ e utilizando-se dados do Censo Agropecuário de 2006. Na análise, considerou-se como produto o valor total da produção e como insumos quantidade de máquinas e implementos agrícolas, área dos estabelecimentos agropecuários e mão-de-obra. Observou-se uma eficiência média de 0,46, ou seja, é possível reduzir 54% dos insumos utilizados sem comprometer a produção. Em geral, os fatores que afetam a eficiência não são os mesmos para todos os municípios e, quando são, não necessariamente tem o mesmo efeito para todos eles. Dentre os municípios menos eficientes (quantis 0,10 e 0,50), o maior percentual de mão-de-obra familiar e de dirigentes com idade inferior à 45 anos associa-se à maior ineficiência.

---

<sup>1</sup> Mestranda em Estatística Aplicada e Biometria, UFV. E-mail: gabriela.franca@ufv.br

<sup>2</sup> Graduanda em Matemática, UFV. E-mail: marcia.guimaraesufv@gmail.com

<sup>3</sup> Professor do Departamento de Estatística, UFV. E-mail: ana.campana@ufv.br; moysesnascim@ufv.br; camila.azevedo@ufv.br

<sup>4</sup> Os autores agradecem à FAPEMIG e CAPES pelo apoio financeiro.



Diferentemente do grupo dos municípios mais eficientes (quantil 0,90), onde um maior percentual de mão-de-obra familiar associa-se à menor ineficiência.

**Palavras-chave:** agricultura, heterogeneidade, produção

## Erros de estimação da tensão de preconsolidação do solo através do ponto de máxima curvatura

Brunna Rithielly Rezende<sup>1 2</sup>

Hygor Amaral Santana<sup>1 2</sup>

Wilhan Valasco dos Santos<sup>1 2</sup>

Anderson Rodrigo da Silva<sup>1 2</sup>

Da curva de compressão do solo é extraído um indicador que expressa a máxima pressão que pode ser aplicada antes da ocorrência de compactação – a tensão de preconsolidação. O método considerado padrão para determinação da tensão é o método de Casagrande, que se baseia na estimativa visual do ponto de máxima curvatura da curva de compressão. No entanto, a subjetividade ligada a essa estimação visual aliada a falta de ajuste dos modelos da curva de compressão podem promover estimativas bastante incoerentes de tensão. O objetivo deste trabalho é eliminar a subjetividade do método de Casagrande e indicar modelo e método de determinação do ponto de máxima curvatura mais adequados para cada formato de curva de compressão uniaxial. Os modelos de van Genuchten e Gompertz foram utilizados para ajustar o índice de vazios em função do logaritmo da tensão aplicada. Utilizou-se três métodos para obtenção de pontos *locais* de máxima curvatura das curvas simuladas: o método das distâncias perpendiculares (PD), o método do platô de resposta linear (LRP) e o método de regressão *spline* bilinear. O modelo de Gompertz é mais facilmente ajustado que o modelo de van Genuchten. O método do platô de resposta linear é drasticamente afetado por alterações no índice de recompressão. O método PD é o menos afetado pelo formato da curva de compressão, apresentando valores intermediários de ponto de máxima curvatura em relação aos métodos *spline* bilinear e LRP, além de apresentar, em geral, menor variabilidade, independente do modelo utilizado.

**Palavras-chave:** *Compactação do solo, Curva de compressão, Simulação.*

---

<sup>1</sup> Instituto Federal Goiano – Campus Urutai. Email: brunnarithielly@hotmail.com

<sup>2</sup> Agradecimentos ao Instituto Federal Goiano – Campus Urutai pelo apoio financeiro.

## Estimating the longitudinal concordance correlation function: an application in fruit color analysis

Thiago de Paula Oliveira <sup>1 2 4</sup>

John Hinde <sup>3 4</sup>

Silvio Sandoval Zocchi <sup>1 4</sup>

The maturity stages of papaya fruit based on peel color is frequently characterized from a sample of four points on the equatorial region measured by a colorimeter. However, this procedure may not be suitable for assessing the papaya's overall mean color and an alternative proposal is to use image acquisition of the whole fruit's peel. Questions of interest are whether a sample on the equatorial region can reproduce a sample over the whole peel region and if the colorimeter can compete with a scanner, or digital camera, in measuring the mean hue over time. The reproducibility can be verified by using the concordance correlation for responses measured on a continuous scale. Thus, in this work we propose a longitudinal concordance correlation (LCC), based on a mixed-effects regression model, to estimate agreement over time among pairs of observations obtained from different combinations between measurement method and sampled peel region. The results show that the papaya's equatorial region is not representative of the whole peel region, suggesting the use of image analysis rather than a colorimeter to measure the mean hue. Moreover, in longitudinal studies the LCC can suggest over which period the two methods are likely to be in agreement and where the simpler colorimeter method could be used. The performance of the LCC is evaluated using a small simulation study.

**KEYWORDS:** Colorimeter, digital image analysis, longitudinal data, mixed-effects model, postharvest

---

<sup>1</sup>Departamento de Ciências Exatas, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo (USP), Av. Pádua Dias, 11, 13418-900 Piracicaba, SP, Brazil

<sup>2</sup>Autor correspondente: *thiago.paula.oliveira@usp.br*

<sup>3</sup>School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, University Rd., Galway, Ireland

<sup>4</sup>Agradecimento ao CNPq e a FAPESP pelo apoio financeiro.

## Estratégias para análise de contagens sub e superdispersas

Eduardo Elias Ribeiro Junior <sup>† 1 2</sup>

Clarice Garcia Borges Demétrio <sup>1</sup>

Na análise de dados em forma de contagens, comumente, a suposição de equidispersão não é adequada e, conseqüentemente, os modelos de regressão Poisson são inapropriados. Importantes avanços na área de análise de contagens têm sido relatados na literatura, principalmente, para modelar diferentes níveis de dispersão, nomeadamente, sub (média > variância) e superdispersão (média < variância). Neste artigo, são revisados os modelos COM-Poisson, *Gamma-Count*, Poisson generalizada e Poisson-Tweedie. A gênese de cada modelo é apresentada, juntamente como um resumo comparativo das distribuições. A similaridade dos modelos COM-Poisson e *Gamma-Count* é destacada, assim como a flexibilidade dos modelos Poisson generalizado e Poisson-Tweedie para modelar superdispersão. A aplicação dos modelos é ilustrada com a análise do número de progênes de *Sitophilus zeamais*, observado em um experimente completamente casualizado com quatro substratos de milho como tratamentos e dez repetições. A implementação computacional é realizada no software R, cujos códigos são disponibilizados em material suplementar.

**Palavras-chave:** *Distribuição COM-Poisson, Distribuição Gamma-Count, Distribuição Família Poisson-Tweedie, Poisson generalizada, Superdispersão, Subdispersão.*

---

<sup>†</sup>Contato: [jreduardo@usp.br](mailto:jreduardo@usp.br)

<sup>1</sup>Departamento de Ciências Exatas (LCE) - ESALQ-USP

<sup>2</sup>Laboratório de Estatística e Geoinformação (LEG) - UFPR

## Estudo Comparativo de três diferentes modelos não lineares para a característica ganho de peso em bovinos de corte, da raça nelore

André Luiz Pinto dos Santos <sup>1 5</sup>

Lêda Valéria Ramos Santana <sup>1 5</sup>

Guilherme Rocha Moreira <sup>2 5</sup>

Frank Sinatra Gomes da Silva <sup>2 5</sup>

Cícero Carlos Ramos de Brito <sup>2 5</sup>

O ajuste de curvas de crescimento na produção animal é uma prática que pode auxiliar na escolha do melhor método de exploração e, conseqüentemente, contribuir para aumentar o lucro do produtor. Neste trabalho, o objetivo foi selecionar o modelo mais adequado ao ajuste das curvas de crescimento em dados de bovinos de corte da raça nelore ao longo dos dias. Para isso, foram utilizados 50 indivíduos em cinco momentos distintos (0, 270, 365, 550, 660 em dias). Para estimar o crescimento em função dos dados de peso-idade, utilizaram-se os modelos não lineares Gompertz, Logístico e Von Bertalanffy. Os parâmetros dos modelos foram estimados pelo método iterativo Gauss-Newton, por meio da técnica dos modelos não-lineares utilizando-se o procedimento nls do programa computacional R. Os parâmetros estimados foram  $\alpha$  (valor assintótico),  $\beta$  (constante de integração, relacionada aos pesos iniciais do animal) e  $k$  (taxa de maturação). O AIC-Critério de Informação de Akaike (AKAIKE 1974) e o BIC-Critério de Informação de Bayes (SCHWARZ 1978), foram utilizados como critérios para a escolha dos modelos que melhor se ajustaram aos dados simulados. Dentre os modelos ajustados, o modelo Von Bertalanffy foi o que apresentou melhor ajuste de acordo com a metodologia e condições em que foi desenvolvido o presente estudo.

**Palavras-chave:** *Curva de crescimento, seleção de modelo, e produção animal.*

---

<sup>1</sup>DEINFO-UFRPE. Doutorando. e-mail: [andredefensor@hotmail.com](mailto:andredefensor@hotmail.com)

<sup>1</sup>DEINFO-UFRPE. Doutoranda.

<sup>2</sup>DEINFO-UFRPE. Professor do departamento de Estatística e Informática.

<sup>2</sup>IFPE. Professor do Instituto Federal de Pernambuco.

<sup>5</sup>Agradecimento ao CNPq pelo apoio financeiro.

---

## **Estudo das espécies Repolho Chato do Quintal e Coração de Boi utilizando teste de hipóteses no campus Monte Carmelo da Universidade Federal de Uberlândia - UFU**

**Maria Thereza Sara Silvestre<sup>1</sup>**

**Claudemir Oliveira Diniz<sup>1</sup>**

**Jair Rocha do Prado<sup>2</sup>**

**Igor Forigo Beloti<sup>3</sup>**

O repolho é uma das hortaliças mais comuns entre as consumidas, apesar de ter o sabor um tanto quanto exótico, é bastante produzida e benéfica. Existem diversas espécies de repolho, entretanto, duas delas chama mais atenção. Foi feita então, a análise sobre essas duas espécies específicas de repolho, sendo elas: Repolho Chato do Quintal e Repolho Coração de Boi. Este experimento faz parte de uma atividade de iniciação científica júnior da UFU. O objetivo da análise foi verificar qual espécie possui melhores características no período de germinação. O experimento foi realizado no Laboratório de Análise de Sementes e Recursos Genéticos - LAGEN, da Universidade Federal de Uberlândia, Campus Monte Carmelo, onde foram alocadas sementes separadamente de cada espécie em gerbox e colocadas na geladeira para germinar. Depois de germinadas, passados sete dias, foi analisado o comprimento das radículas, hipocótilo, comprimento total e massa verde das plântulas. Logo em seguida, foi colocado o material em uma estufa e analisado a massa seca. Após essas análises, foram feitos testes de hipóteses com uso do software R, para comparação de médias e variâncias considerando as variáveis e as espécies de repolho mencionadas anteriormente. Concluiu-se, por meio dos testes de hipóteses, ao nível de significância de 5%, que as médias das duas espécies de repolho diferem estatisticamente para as variáveis: radícula, hipocótilo, proporção de sementes germinadas, massa seca e comprimento total. Enquanto que os valores de massa verde foram considerados iguais.

**Palavras-chave:** *Plântula, Experimento, Repolho, Teste de hipóteses.*

---

<sup>1</sup> Escola Estadual Professor Vicente Lopes Perez, Monte Carmelo – MG.

<sup>2</sup> FAMAT – UFU

<sup>3</sup> ICIAG - UFU

## **Estudo das Séries Temporais das temperaturas média, máxima e mínima para o município de Curitibanos/SC**

**Nicole Orsi<sup>1</sup>**

**Crysttian Arantes Paixão<sup>2</sup>**

**Leosane Cristina Bosco<sup>3</sup>**

Fatores climáticos, como a temperatura, têm impacto em diversos processos em ambientes urbanos e rurais. Uma forma de evitar danos provenientes de eventos climáticos extremos é realizar a previsão destes, tornando possível a realização de um planejamento prévio. A previsão em Séries Temporais é feita pela análise de dados antecedentes, avaliando componentes das séries como sazonalidade, tendência e estacionariedade. O objetivo deste trabalho foi realizar a previsão das temperaturas média, máxima e mínima para o município de Curitibanos, em Santa Catarina, para os anos de 2018 até 2020, utilizando Séries Temporais. A análise e previsão das séries temporais das temperaturas foi realizada utilizando o software R. A previsão foi estimada por meio dos Modelos Autorregressivos Integrados de Médias Móveis (ARIMA). Portanto, a partir da construção deste modelo, foi possível realizar a previsão das temperaturas média, máxima e mínima para um período de três anos. De acordo com os testes executados e a previsão observou-se que as séries temporais para as temperaturas média, máxima e mínima não apresentaram tendência, mas apresentaram estacionariedade e sazonalidade. Portanto, cada série se desenvolve no tempo aleatoriamente ao redor de uma média constante e os fenômenos se repetem durante períodos similares de tempo. Pôde-se perceber que a série temporal das temperaturas, para o município de Curitibanos/SC, possui um comportamento bastante definido ao longo dos anos.

**Palavras-chave:** *Previsão, Séries Temporais, Temperatura.*

---

<sup>1</sup> Acadêmica de Agronomia – Universidade Federal de Santa Catarina (UFSC), Campus de Curitibanos. E-mail: nicoleorsi11@gmail.com.

<sup>2</sup> Professor Doutor – Universidade Federal de Santa Catarina (UFSC), Campus de Curitibanos, Centro de Ciências Rurais, Departamento de Ciências Naturais e Sociais.

<sup>3</sup> Professora Doutora – Universidade Federal de Santa Catarina (UFSC), Campus de Curitibanos, Centro de Ciências Rurais, Departamento de Agricultura, Biodiversidade e Florestas.

## **Estudo do desenvolvimento inicial de plântulas de almeirão Radiche e Pão de açúcar com o uso de testes de hipóteses**

**Claudemir Oliveira Diniz<sup>1</sup>**

**Maria Thereza Sara Silvestre<sup>1</sup>**

**Jair Rocha do Prado<sup>2</sup>**

**Igor Forigo Beloti<sup>3</sup>**

O trabalho desenvolvido a seguir, faz parte de uma atividade de iniciação científica júnior da UFU, onde seu principal intuito foi a comparação de duas espécies de almeirão usando testes de hipóteses. O estudo foi feito a partir de plântulas das espécies de almeirão “Pão de Açúcar” e “Radiche”. O experimento foi realizado no Laboratório de Análise de Sementes e Recursos Genéticos - LAGEN, da Universidade Federal de Uberlândia, Campus Monte Carmelo. Inicialmente, foram depositadas sementes em recipientes adequados e em condições propícias, para que pudessem germinar e assim fornecer os resultados a serem discutidos. Dos exemplares de plântulas obtidos, mediram-se os comprimentos: totais, das radículas e dos hipocótilos de cada espécie, além da pesagem da massa verde e da massa seca. A partir dos dados coletados foi possível fazer as comparações de médias e variâncias, considerando cada variável mencionada e cada variedade de almeirão. Para este estudo utilizou-se a metodologia de testes de hipóteses, sendo que as análises foram realizadas por meio do software R. Pôde-se concluir a partir dos resultados obtidos nos testes, a um nível de significância de 5%, a existência de diferença nas médias das espécies, para os valores relativos às variáveis: comprimento radicular, comprimento dos hipocótilos, massa verde e comprimento total. A única variável que apresentou certa similaridade quanto aos resultados das médias obtidas foi a massa seca.

**Palavras-chave:** *Plântula, Experimento, Teste de hipóteses, Almeirão.*

---

<sup>1</sup> Escola Estadual Professor Vicente Lopes Perez, Monte Carmelo – MG.

<sup>2</sup> FAMAT – UFU

<sup>3</sup> ICIAG - UFU



## Estudo do Endurecimento e Escurecimento de Feijão utilizando Componentes Principais Esparsas

Flávia Adriane Pestana de Oliveira <sup>1</sup>

Dr. George Freitas von Borries <sup>2</sup>

Dra. Priscila Zaczuk Bassinello <sup>3</sup>

Dra. Anna Cristina Lanna <sup>3</sup>

Ma. Rosangela Nunes Carvalho <sup>3</sup>

O feijão passa por modificações no tegumento e endurecimento que podem levar a perda de qualidade do grão. Com o intuito de controlar estas perdas de qualidade foram analisados os fatores de tempo de armazenamento (5 níveis), temperatura (baixa, média e alta) e genótipo (5 materiais com 3 repetições cada). Existem indicações de que estes fatores podem afetar as variações das medidas referentes ao escurecimento e endurecimento do feijão. A aplicação da Análise de Componentes Principais (ACP) tradicional para redução dos dados, apresenta problemas devido ao grande número de variáveis em relação ao tamanho da amostra. Uma alternativa é a aplicação de ACP Esparsa. Esta abordagem mostrou a existência de variação considerável em certos grupos de variáveis e relações dentro e entre os grupos de escurecimento e endurecimento, para alguns dos genótipos. Porém, não foram encontradas relações com o dia e a temperatura de coleta das medidas.

**Palavras-chave:** *Componentes Principais Esparsas, Feijão, Medidas de endurecimento, Medidas de escurecimento.*

---

<sup>1</sup>Bacharela em Estatística, Universidade de Brasília. E-mail: *flavidriane@hotmail.com*

<sup>2</sup>Professor, Departamento de Estatística, Universidade de Brasília, e-mail: *gborries@unb.br*

<sup>3</sup>Pesquisadora, EMBRAPA Rice and Beans.

<sup>3</sup>Pesquisadora, EMBRAPA Rice and Beans.

<sup>3</sup>Analista, EMBRAPA Rice and Beans.

## Estudos de simulação da relação entre modelos logit e biplots

Haiany Aparecida Ferreira <sup>1 4</sup>

Carla Regina Guimarães Brighentti <sup>2 4</sup>

Marcelo Ângelo Cirillo <sup>3 4</sup>

Um método estatístico que vem sendo muito utilizado em diversas áreas atualmente é a técnica multivariada Biplot. Esta técnica consiste na redução de dimensão e pode ser aplicada, por exemplo, tanto na análise de componentes principais quanto de correspondência. Sendo assim, o objetivo deste trabalho consiste em propor um estudo computacional utilizando a simulação Monte Carlo, referente ao estudo do efeito da heterogeneidade entre tabelas de frequências diagnosticada por meio de modelos logit binomial e sua relação com os biplots. Para a aplicação desta técnica foi utilizada a teoria da decomposição de valores singulares em cenários com diferentes graus de correlação. Além disso, outro fator avaliado levado em consideração foram as somas de quadrados em cada cenário. Deste modo, a seleção dos cenários é feita de acordo com os resultados e interpretações gráficas obtidas mais plausíveis.

**Palavras-chave:** *Biplot, Logit Binomial, Decomposição de Valores Singulares.*

---

<sup>1</sup>DES - UFLA. e-mail: [haianyferreira@yahoo.com.br](mailto:haianyferreira@yahoo.com.br)

<sup>2</sup>DZOO - UFSJ.

<sup>3</sup>DES - UFLA.

<sup>4</sup>Agradecimento ao CNPq e Fapemig pelo apoio financeiro.

## Imputação de dados e análise de variância: uma aplicação em dados reais

Elisandra Lúcia Moro Stochero<sup>1</sup>

Luciane Flores Jacobi<sup>2</sup>

Alessandro Dal'Col Lúcio<sup>3</sup>

Mateus Bergamo<sup>4</sup>

Otávio Flores Jacobi<sup>5</sup>

Existem técnicas que buscam minimizar o viés de resultados quando ocorre perda de dados durante o desenvolvimento de um experimento. Estas técnicas são conhecidas como métodos de imputação de dados faltantes, as quais consistem em determinar estimativas para os valores que foram perdidos com o objetivo de obter resultados mais precisos. A proposta desta pesquisa, que se encontra em andamento, é avaliar se há diferença nos resultados da Análise de variância quando aplicado métodos de imputação em um banco de dados reais. O banco de dados inicial, completo, disponibilizado para o estudo trata de um experimento com tomate tipo salada, híbrido *Grandeur*, conduzido no Departamento de Fitotecnia da Universidade Federal de Santa Maria, neste experimento foi utilizado o delineamento blocos ao acaso com três repetições e 12 tratamentos. Serão removidos 10% dos dados iniciais gerando um novo banco de dados incompleto, desbalanceado. Através de diferentes métodos de imputação serão obtidas estimativas para os dados que serão retirados e incluídas no banco de dados incompleto, gerando, assim, um terceiro banco de dados, completo, obtido a partir da imputação. Em cada uma das etapas será realizada a ANOVA e posteriormente serão comparados os resultados a fim de verificar se a partir da aplicação do método de imputação, obtém-se resultados com menor viés na ANOVA.

**Palavras-chave:** *ANOVA, Dados faltantes, Métodos de Imputação, Planejamento de experimentos, Solanum lycopersicum.*

---

<sup>1</sup> BACHARELADO EM ESTATÍSTICA – UFSM. e-mail elismoro2016@gmail.com  
<sup>2</sup> DEPARTAMENTO DE ESTATÍSTICA – UFSM  
<sup>3</sup> DEPARTAMENTO DE FITOTECNIA – UFSM  
<sup>4</sup> UFSM  
<sup>5</sup> UFRGS

## Influência da magnetita (Fe<sub>3</sub>O<sub>4</sub>) no processo de absorção de nutrientes em alface crespa Vanda (*Lactuca sativa*)

Clara Augusta Marques Natalin<sup>1 2</sup>

Amanda Liz Pacífico Manfrim Perticarrari<sup>2 2</sup>

Priscila Lupino Grato<sup>2 2</sup>

Luis Fabiano Palaretti<sup>2 2</sup>

José Marques Junior<sup>2 2</sup>

**Palavras-chave:** *Alface, Crescimento vegetativo, Magnetita.*

### Introdução

No Brasil grande parte dos solos são muito intemperizados, ácidos e de baixa fertilidade natural [1]. A exploração agrícola continuada associada ao manejo inadequado acarreta a degradação química do solo, limitando o desenvolvimento e o potencial agrícola das culturas. Uma das formas de remediar a baixa fertilidade é com a aplicação de magnetita, sendo essa uma pedra natural com alto teor de ferro [2].

Esse tratamento pode afetar a produção de fitormônio, melhorando as células, o crescimento de atividade e da planta [3]. [4] relatou que com o aumento dos níveis de magnetita, aumentou incrementos graduais em N, P, K e nas concentrações de Fe na couve-flor. Resultados semelhantes foram obtido por [5] que descobriram que as maiores concentrações de Fe nas folhas de Thompson Seedless<sup>3</sup> foi associada com árvores cultivadas em solo tratado com magnetita. Visando aumentar o potencial agrícola, têm sido proposto o uso dela.

### Objetivo

Avaliar o efeito da utilização de magnetita no cultivo de alface crespa Vanda (*Lactuca sativa*), visando a dosagem ótima de magnetita na absorção de nutrientes dessa variedade.

---

1 AGRO – UNESP – FCAV. e-mail claranatalin@grad.fcav.unesp.br

2 DCE – UNESP – FCAV

3 Considerada a uva sem semente de maior importância no mundo [6]

## Material e métodos

O experimento e as análises<sup>4</sup> serão realizados na UNESP<sup>5</sup>. Serão utilizados vasos leonard, preenchidos com areia esterilizada e diferentes concentrações de magnetita. Para atender as necessidades nutricionais da cultura será utilizado solução nutritiva<sup>6</sup>.

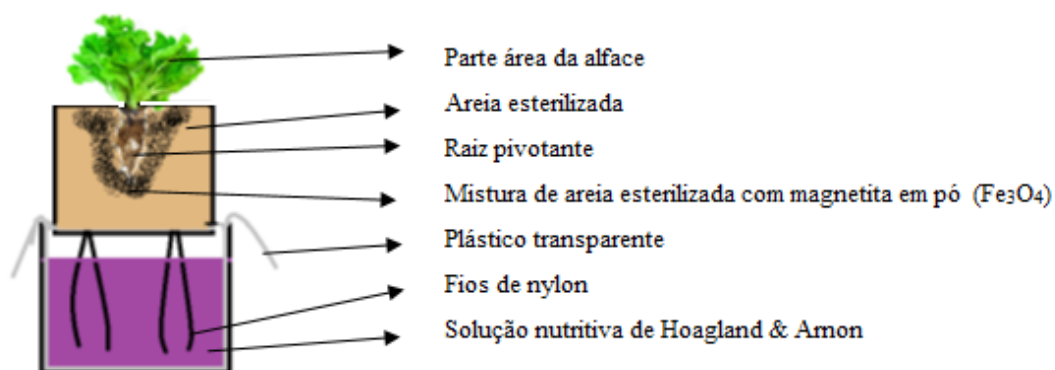


FIGURA 1. Esquema do vaso leonard

## Resultados e discussões

Com a utilização da magnetita se espera ter um aumento na absorção dos nutrientes e o consequente aumento da área vegetativa.

## Referências Bibliográficas

- [1] EMBRAPA. Sistema brasileiro de classificação de solos. Rio de Janeiro: Embrapa Solos, 412p, 1999.
- [2] MANSOUR, E. R. Effect of some culture practices on cauliflower tolerance to salinity under Ras Suder conditions. Msr Thesis. Fac. of Agric. **Horticulture Dept. Ain Shams Univ**, 2007.

---

4 Os dados coletados serão analisados utilizando Testes Estatísticos e o software R para observar os efeitos da magnetita na absorção de nutrientes em alface crespa Vanda (*Lactuca sativa*)

5 No campus de Jaboticabal/SP, sendo conduzido em casa de vegetação

6 De Hoagland & Arnon

[3] MAHESHWARI, Basant L.; GREWAL, Harsharn Singh. Magnetic treatment of irrigation water: Its effects on vegetable crop yield and water productivity. *Agricultural water management*, v. 96, n. 8, p. 1229-1236, 2009.

[4] ALI, T. B.; KHALIL, S. E.; KHALIL, A. M. Magnetic treatments of *Capsicum annuum* L. grown under saline irrigation conditions. *J. Appl. Sci. Res*, v. 7, n. 11, p. 1558-68, 2011.

[5] ISMAIL, A. E. et al. Effect of magnetic iron ore, metal compound fertilizer and bio-NK in controlling root-knot nematode of grapevine in a newly reclaimed area of Egypt. *Pakistan Journal of Nematology*, v. 28, n. 2, p. 307-328, 2010.

[6] EMBRAPA; **Thompson Seedless**. Disponível em: [http://www.agencia.cnptia.embrapa.br/gestor/uva\\_de\\_mesa/arvore/CONT000gn4xq0u202wx5ok0liq1mqosbxgyy.html#](http://www.agencia.cnptia.embrapa.br/gestor/uva_de_mesa/arvore/CONT000gn4xq0u202wx5ok0liq1mqosbxgyy.html#). Acessado em 10 de março de 2018.

---

## MÉTODO DE SELEÇÃO DE PREDITORES LINEARES GEOESTATÍSTICOS VIA ABORDAGEM DO CAMPO ALEATÓRIO GAUSSIANO

Peter de Matos Campos<sup>1</sup>

Marcelo Silva de Oliveira<sup>2</sup>

**Resumo** – O problema da predição sob dependência espacial continua a figurar como um grande desafio devido à complexidade dos fenômenos estudados, grande disponibilidade de informação auxiliar, elevado número de métodos de interpolação espacial e, também, pela incipiência dos critérios para a escolha do método de interpolação espacial mais adequado para cada situação. Quando o estudo é univariado a resposta é: krigagem, pois seus preditores são BLUP. Quando variáveis auxiliares são utilizadas não se encontra uma resposta consensual na literatura. Provavelmente por essas razões, técnicas alternativas têm sido adotadas por pesquisadores para incluir variáveis auxiliares em seus estudos, já que tais técnicas alternativas podem seguir uma lógica prática, entretanto, não garantem que seus preditores sejam um BLUP. Assim, o objetivo deste trabalho é propor um critério teórico para a seleção de preditores de krigagem linear que permaneçam sendo BLUP. Este critério é tratado neste estudo utilizando a abordagem do campo aleatório gaussiano (ACAG). Para a definição destes critérios, seis cenários decorrentes da combinação da partição do erro devido à dependência espacial  $\varepsilon'(s_0)$  e pela existência ou não de autocorrelação espacial (AC), correlação cruzada espacial (CC) e correlação simples não espacial (CS), sob um CAG foram estudados, comprovando-se a validade desses critérios teóricos. Também um critério prático baseado na classificação de um índice de dependência espacial *IDE*(%) encontrado na literatura define para quais valores ou faixa de valores a AC, CC e CS são fortes, possibilitando a indicação prática do melhor preditor de krigagem linear, especialmente quando variáveis auxiliares são introduzidas no estudo.

**Palavras-chave:** Correlações. Gaussianidade local. Índice de dependência espacial. Krigagem. Variáveis auxiliares.

---

<sup>1</sup> DEMAT – UFSJ, e-mail: pmcampos@ufs.br.

<sup>2</sup> DES – UFLA, e-mail : marcelo.oliveira@des.ufla.br.

# Modelagem da temperatura mínima mensal de Piracicaba-SP via Distribuição Generalizada de Valores Extremos

Andréia do Carmo de Oliveira<sup>1 2 4</sup>

Fabricio Goecking Avelar<sup>3 4</sup>

A ocorrência de eventos climáticos extremos, como por exemplo, precipitações extremas, velocidades extremas de ventos e temperaturas extremas podem gerar impactos em diversas atividades, como a saúde pública e agricultura. A previsão da ocorrência e da intensidade desses fenômenos é importante para o planejamento de ações dos órgãos públicos e população em geral. O presente trabalho teve como objetivo estudar o comportamento da temperatura mínima mensal, dos meses de maio a setembro, de Piracicaba-SP. Os dados diários de temperatura mínima foram obtidos junto à Estação Convencional do Posto Agrometeorológico da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP), localizada em Piracicaba-SP, no período de 1917 a 2017. A série de dados mensal de temperatura mínima foi obtida tomando-se a menor observação de cada mês. O teste Ljung-Box foi utilizado para verificar a independência das séries de temperatura mínima mensal e, o teste de Mann-Kendall, para analisar a presença de tendência. Ao nível de 5% de significância, os dados de temperatura mínima mensais foram considerados independentes e não estacionário para todos os meses estudados. A Distribuição Generalizada de Valores Extremos com tendência linear no parâmetro posição foi ajustada aos dados e a qualidade dos ajustes foi verificada pela análise dos gráficos Quantil-Quantil (QQ-Plot) e Probabilidade-Probabilidade (PP-Plot). Foram calculados os níveis de retorno para os tempo de retorno de 5, 10 e 20 anos. Concluiu-se que, a partir do ano de 1976 a temperatura mínima de Piracicaba-SP está aumentando.

**Palavras-chave:** *Tendência, Eventos extremos, Estimador de Máxima Verossimilhança.*

---

<sup>1</sup>Discente de Licenciatura em Matemática - UNIFAL-MG. e-mail: [andreiaoliveira.math@gmail.com](mailto:andreiaoliveira.math@gmail.com)

<sup>2</sup>Bolsista FAPEMIG

<sup>3</sup>ICEX - UNIFAL-MG. email: [fabricio@unifal-mg.edu.br](mailto:fabricio@unifal-mg.edu.br)

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.



## Modelagem do acúmulo de massa seca em pastagem de aveia e azevém, em cultivo estreme ou consorciados, sob pastejo

Vanessa Ruiz Favaro<sup>1</sup>

Simone Silmara Werner<sup>1</sup>

Ulisses de Arruda Cordova<sup>1</sup>

Maicon Gaissler Lorena Pinto<sup>1</sup>

O presente trabalho objetivou descrever o acúmulo de massa seca (kg/ha) durante o ciclo produtivo de pastagens de aveia branca (cv. Milton), azevém anual (cv. Winter Star) e o consórcio dessas forrageiras em sistema rotacionado em função de variáveis climáticas. As pastagens foram implantadas na Epagri/Estação Experimental de Lages (EEL) durante 3 safras consecutivas em uma área de 0,4 hectares cada. Para estimativa da matéria seca foram feitos cinco cortes, rente ao solo com quadro amostral de 0,5 x 0,5m, em cada piquete. A entrada e saída dos animais foi determinada com base na recomendação técnica para altura de manejo das espécies avaliadas. Os dados climáticos foram obtidos da estação climatológica localizada na EEL. Para análise dos dados utilizou-se o modelo normal considerando como variáveis explicativas para a média os valores acumulados médios de precipitação, temperatura máxima, média e mínima no período, os efeitos das safras, espécies e dias após a implantação das pastagens. Para descrever a variância foram considerados os efeitos de safra, espécies e dias após a implantação. A seleção dos modelos foi realizada utilizando o critério de informação de Akaike (AIC) e análise de diagnóstico por meio dos gráficos de resíduos. Todas as análises foram realizadas utilizando o ambiente R. A variância foi maior na primeira safra. Observou-se relação entre as variáveis climáticas chuva acumulada e temperatura mínima com o acúmulo da massa seca. Não houve diferença entre os parâmetros estimados para a aveia e o consórcio aveia + azevém nas diferentes safras. O coeficiente angular associado a dias na terceira safra para o azevém foi superior ao estimado na primeira safra. Com base na análise de resíduos, o modelo utilizado foi satisfatório para explicar o acúmulo de massa seca.

**Palavras-chave:** *Heterogeneidade de variâncias, variáveis climáticas, ambiente R.*

---

<sup>1</sup>Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina - Epagri/Estação Experimental de Lages. e-mail: [vanessafavaro@epagri.sc.gov.br](mailto:vanessafavaro@epagri.sc.gov.br), [simone Werner@epagri.sc.gov.br](mailto:simone Werner@epagri.sc.gov.br), [ulisses@epagri.sc.gov.br](mailto:ulisses@epagri.sc.gov.br), [maiconpinto@epagri.sc.gov.br](mailto:maiconpinto@epagri.sc.gov.br)

---

## MODELAGEM DO CRESCIMENTO DE POEDEIRAS COMERCIAIS, CORPO E DE SEUS COMPONENTES, A PARTIR DE OBSERVAÇÕES "in vivo"

Malheiros, E. B.

Alves Junior, W.

Sakomura, N. K.

Alves, E. P

O modelo de crescimento de aves de Emmans (1981), descreve o animal, por métodos destrutivos. O DXA (absorciometria de raios-X) é um método alternativo que permite, a partir da composição corporal *in vivo*, estudar variações do crescimento animal (SCHOLZ et al., 2015). Ensaio realizado: 90 aves - Dekalb Brown, avaliações: 1-126 dias.

Elaborou-se modelos de predição da composição corporal *in vivo*, a partir de das variáveis do DXA: X1:MT - Massa total, X2:MG - Massa gorda, X3:MM - massa magra (água + proteínas), X4:CMO - conteúdo mineral ósseo e, para avaliar quais delas interferem no peso, usou o modelo:  $\ln(Y) = B_0 + B_1 \ln(X_1) + B_2 \ln(X_2) + B_3 \ln(X_3) + B_4 \ln(X_4) + \varepsilon$ ; Y - cada variável observada "in vivo": PCLP - peso do corpo livre de penas, MM - matéria mineral, GD - massa de gordura, PB - teor de proteína bruta e PACLP - peso de água do corpo livre de penas. Pelo método "stepwise" selecionou-se os modelos: PCLP = f(MT); MM = f(CMO, MM); GD = f(CMO, MG, MM, MT); PB = f(MM) e PACLP = f(MM, MT).

Com os valores estimados, ajustou-se o modelo de Gompertz,

$$W_i = W_m \times \exp(-\exp(\ln(-\ln(W_i / W_m)) - (B \times t)))$$

onde:  $W_i$ =peso-inicial,  $W_m$ =peso-maturidade, B = taxa crescimento,  $t^*$  - ponto de inflexão.

A partir deste modelo obteve-se as estimativas dos parâmetros:  $W_i(g)$ :  $32,27 \pm 0,57$ ;  $W_m(g)$ :  $1808,70 \pm 56,57$ ; B(/dia):  $0,0243 \pm 0,00007$ ;  $t^*$  (dias):  $56,57 \pm 0,90$ .

**Palavras-chave:** Modelagem de crescimento, poedeiras comerciais; método não destrutivo.

## Modelando curvas de crescimento animal com os modelos aditivos generalizados para locação, escala e forma

Luiz R. Nakamura <sup>1</sup>

Thiago G. Ramires <sup>2</sup>

Ana J. Righetto <sup>3</sup>

Rodrigo R. Pescim <sup>4</sup>

Taciana V. Savian <sup>5</sup>

Curvas de crescimento (CC) animal (peso×idades), em geral, são analisadas por meio de modelos não-lineares devido à interpretação prática de seus parâmetros. Entretanto, em determinadas situações, como o caso de CCs que apresentam comportamentos cíclicos, isto é, multifásicos, a convergência de tais parâmetros não é trivial. Neste sentido, propõe-se neste trabalho, a utilização dos modelos aditivos generalizados para locação, escala e forma (GAMLSS) como uma alternativa para o ajuste de CCs. Para tal, foi utilizado um conjunto de dados composto por 55 fêmeas da raça Hereford de uma fazenda pertencente ao município de Bagé-RS, nascidas entre 1999 e 2001. A coleta dos pesos dos animais foi realizada de 15 em 15 dias, do seu nascimento até os 615 dias e apresenta o comportamento multifásico. Baseado nos modelos GAMLSS, foi observado que a média de peso dos animais aumenta de seu nascimento até os 200 dias, depois decresce dos 200 aos 400 dias em uma taxa mais devagar, cresce dos 400 aos 500 dias e, finalmente, o peso permanece praticamente constante até os 615 dias. Esse comportamento da média pode ser explicado devido a, por exemplo, o excesso de chuvas ou a escassez de comida. Com relação à variabilidade dos dados, pode-se dizer que, como esperado, a variância dos dados é muito menor em animais mais novos e aumenta conforme o envelhecimento dos mesmos.

**Palavras-chave:** *Funções de suavização, GAMLSS, Modelos não-lineares.*

---

<sup>1</sup>INE - UFSC. e-mail: [luiz.nakamura@ufsc.br](mailto:luiz.nakamura@ufsc.br)

<sup>2</sup>DAMAT - UTFPR.

<sup>3</sup>IAPAR.

<sup>4</sup>DSTA - UEL.

<sup>5</sup>LCE - USP/ESALQ.

## Modelo AMMI no estudo de comportamento ingestivo de novilhas

Cristiano Nunes Nesi <sup>1</sup>

Felipe Jochims <sup>2</sup>

O objetivo deste trabalho foi aplicar o modelo AMMI para estudar a interação entre ambiente e atividades do comportamento ingestivo em novilhas. Avaliações foram realizadas visualmente a cada 10 minutos e por 24 horas consecutivas em seis áreas (repetições), uma vez em cada estação do ano. Mediu-se o tempo total de pastejo, de ruminação e de ócio em cada turno em quatro animais por repetição. Considerou-se como ambiente a combinação entre estação do ano e turno do dia (manhã, tarde e noite), o que resultou em 12 ambientes (por exemplo, verão/manhã, verão/tarde e verão/noite). Empregou-se o modelo AMMI que modela de efeitos principais aditivos e interação multiplicativa, representado num biplot. Utilizou-se o programa R e a função AMMI do pacote *agricolae*. O efeito significativo de atividade indicou tempos diferentes para cada atividade. Para ambientes, o efeito significativo mostra que o estudo foi realizado em ambientes diferentes. A interação significativa indica que ambiente e atividade, separadamente, não explicam toda a variação no tempo, resultando em diferentes performances das atividades nas condições avaliadas. Ruminação interage positivamente com as noites da primavera, verão e outono, enquanto ócio interage positivamente com a noite e manhã de inverno. Pastejo interage positivamente com as tardes de todas as estações. As manhãs de primavera e verão contribuem pouco para a interação. As atividades de ócio e ruminação estiveram abaixo da média geral, evidenciando que a atividade principal dos animais em pastejo é a procura e apreensão de alimento, o que é influenciado pelo ambiente, principalmente luminosidade e temperatura.

**Palavras-chave:** *Comportamento animal, Interação, Pastagens, Biplot.*

---

<sup>1</sup>Epagri/Cepaf; UNOESC/Xanxerê. e-mail: [cristiano@epagri.sc.gov.br](mailto:cristiano@epagri.sc.gov.br)

<sup>2</sup>Epagri/Cepaf. e-mail: [felipejochims@epagri.sc.gov.br](mailto:felipejochims@epagri.sc.gov.br)

# Modelo logístico com efeito aleatório para descrever o progresso temporal do cancro cítrico

Cristiano Nunes Nesi <sup>1</sup>

Simone Silmara Werner <sup>2</sup>

Eduardo Cesar Brugnara <sup>3</sup>

A proposta deste trabalho foi avaliar a aplicação do modelo logístico com efeitos aleatórios para descrever o progresso temporal da incidência de cancro cítrico nas folhas de laranjeiras em função da posição do ramo na copa. Os dados foram obtidos por amostragens quinzenais, durante o primeiro fluxo de brotação da estação. Em cada avaliação foram aleatoriamente selecionadas 20 plantas, das quais se avaliou dois ramos em crescimento por quadrante (Norte, Sul, Leste e Oeste). A incidência de cancro cítrico foi obtida pelo quociente entre o número de folhas com sintomas e o número total de folhas do ramo. Para descrever o progresso da incidência em função do tempo para cada quadrante na planta ajustou-se o modelo logístico  $y = \frac{\beta_1}{1 + e^{\frac{\beta_2 - x}{\beta_3}}}$ . Assumindo a inclusão do efeito aleatório de planta em  $\beta_1$ ,  $\beta_2$  e  $\beta_3$ , o modelo escrito com efeitos aleatórios é dado por:  $y_{jik} = \frac{(\beta_{1i} + b_{1ij})}{1 + e^{\frac{((\beta_{2i} + b_{2ij}) - x_k)}{(\beta_{3i} + b_{3ij})}}} + \epsilon_{ijk}$  em que  $y_{ijk}$  representa a  $k$ -ésima mensuração obtida no  $j$ -ésimo ramo para a  $i$ -ésima planta. Assume-se que o vetor  $b_{ij}$  de todos os efeitos aleatórios é proveniente de uma distribuição normal,  $b_{ij} \sim N(0, (D)_{ij})$ , em que  $D$  representa a matriz de variâncias e covariâncias associada aos efeitos aleatórios e que observações feitas em diferentes parcelas são independentes entre si, ou seja,  $\epsilon_{ij} \sim N(0, \sigma^2 I)$ . Os parâmetros do modelo foram estimados pelo método da máxima verossimilhança. Para seleção do modelo mais adequado aos dados utilizou-se o critério de informação de Akaike - AIC e para verificação do ajuste os gráficos de resíduos. Independente do quadrante avaliado, o modelo com efeito aleatório associado apenas ao parâmetro  $\beta_1$  (assíntota) apresentou melhor ajuste. Os quadrantes norte e oeste foram os que apresentaram os maiores valores para o parâmetro  $\beta_1$  evidenciando uma maior incidência do cancro. A utilização do modelo logístico com efeitos aleatórios possibilitou o ajuste da heterogeneidade de variâncias presente e a comparação entre os quatro quadrantes avaliados.

**Palavras-chave:** *heterogeneidade de variâncias, modelos não lineares, dados longitudinais.*

<sup>1</sup>Epagri/Cepaf; UNOESC/Xanxerê. e-mail: [cristiano@epagri.sc.gov.br](mailto:cristiano@epagri.sc.gov.br)

<sup>2</sup>Epagri/EE Lages. e-mail: [simonewerner@epagri.sc.gov.br](mailto:simonewerner@epagri.sc.gov.br)

<sup>3</sup>Epagri/Cepaf. e-mail: [eduardobrugnara@epagri.sc.gov.br](mailto:eduardobrugnara@epagri.sc.gov.br)

## Modelos de sobrevivência para dados grupados da ação de repelentes em abelhas *Trigona spinipes*

Luiz Fernando Silva Resende<sup>1 3</sup>

Carla Regina Guimarães Brighenti<sup>1 2 3</sup>

Deodoro Magno Brighenti dos Santos<sup>2 3</sup>

Juliana Rodrigues Alves<sup>2</sup>

Isabela da Silva Lima<sup>2 3</sup>

Uma espécie de abelha sem ferrão que tem causado danos a flor copo-de-leite é a abelha-irapuá (*Trigona spinipes*). Ao coletar o pólen elas danificam as inflorescências da flor deixando-a inviável para comercialização. Para evitar este dano uma alternativa é aplicar um repelente natural que não ocasione mortalidade às abelhas, apesar do contato tanto durante a aplicação do produto por pulverização, quanto por ingestão ao coletar o pólen. Propõe-se neste trabalho o ajuste de modelos para dados agrupados na avaliação da sobrevivência de abelhas *T. spinipes*, submetidas a diferentes formas de aplicação de repelentes naturais. Para avaliação do número de abelhas mortas, os dados foram coletados a cada 12 horas, sendo a variável resposta o tempo até a ocorrência da morte da abelha e o método de aplicação do repelente considerado como covariável. Foi realizada a análise não paramétrica utilizando estimador limite-produto modificado, o ajuste pelo modelo paramétrico de Weibull e a técnica semiparamétrica por modelos de regressão discretos. As análises foram realizadas utilizando o pacote icenReg no R. Pode-se dizer que a ingestão do repelente causa menor dano às abelhas-irapuá do que a pulverização do mesmo, principalmente nas primeiras 288 horas. O melhor ajuste foi obtido pelo modelo semiparamétrico, com valores bem próximos aos obtidos pela estimativa de Turnbull. Já no modelo paramétrico de Weibull, observou-se muitos valores fora do intervalo de confiança.

**Palavras-chave:** *Trigona spinipe*, sobrevivência, método de aplicação, ingestão, pulverização.

1 DES – UFLA. luizfsr1@hotmail.com

2 DEZOO – UFSJ/PGEE-UFLA. carlabrighenti@ufsj.edu.br

3 Agradecimento ao CNPq e a FAPEMIG pelo apoio financeiro.

## Modelos não-lineares para descrever a curva de crescimento de ovinos da raça Santa Inês

Rayane Santos Leite <sup>1 5</sup>

André Luiz Pinto dos Santos <sup>1 5</sup>

Elielma Santana de Jesus <sup>2 5</sup>

Fernando Henrique Antunes de Araújo <sup>2 5</sup>

Guilherme Rocha Moreira <sup>3 5</sup>

Neste trabalho, o objetivo foi selecionar o modelo mais adequado ao ajuste das curvas de crescimento em dados de ovinos da raça Santa Inês ao longo dos dias. Para isso, foram utilizados 100 indivíduos em cinco momentos distintos (0, 28, 56, 84, 112, 140, 168, 196) dias. Para estimar o crescimento em função dos dados de peso-idade, utilizaram-se os modelos não lineares Gompertz, Logístico, Brody e Von Bertalanffy. Os parâmetros dos modelos foram estimados pelo método iterativo Gauss-Newton, por meio da técnica dos modelos não-lineares utilizando-se o procedimento nls do programa computacional R. Os parâmetros estimados foram  $\alpha$  (valor assintótico),  $\beta$  (constante de integração) e  $k$  (taxa de maturação). O AIC (Informação de Akaike), BIC (Informação de Bayes) e o quadrado médio do resíduo (QMR), foram utilizados como critérios para a escolha dos modelos que melhor se ajustaram aos dados simulados. Para verificação das pressuposições do modelo foi realizado a análise de resíduo. Pelo gráfico do diagrama de dispersão dos resíduos em função dos valores ajustados, o gráfico quantil-quantil normal dos resíduos e pelo teste de Shapiro-Wilk ( $p$ -valor = 0.95) não se têm razões para negar o atendimento dos pressupostos. Dentre os modelos ajustados o modelo Gompertz foi o que apresentou melhor ajuste de acordo com a metodologia e condições em que foi desenvolvido o presente estudo.

**Palavras-chave:** *Produção animal, seleção de modelo, taxa de crescimento.*

---

<sup>1</sup>DEINFO-UFRPE. Mestranda. e-mail: [andredefensor@hotmail.com](mailto:andredefensor@hotmail.com)

<sup>1</sup>DEINFO-UFRPE. Doutorando.

<sup>2</sup>DEINFO-UFRPE. Mestranda.

<sup>2</sup>DEINFO-UFRPE. Doutorando.

<sup>3</sup>DEINFO-UFRPE. Professor do departamento de Estatística e Informática.

## Modelos para predição de área foliar de espécies forrageiras

Hygor Amaral Santana<sup>1 2</sup>

Brunna Rithielly Rezende<sup>1 2</sup>

Wilhan Valasco dos Santos<sup>1 2</sup>

Anderson Rodrigo da Silva<sup>1 2</sup>

A área foliar é uma variável importante em estudos fitofisiológicos, por estar relacionada com a taxa fotossintética e outros processos. Assim, surge a necessidade de métodos rápidos e precisos para estimar a área foliar. O objetivo do trabalho foi ajustar modelos de regressão para predizer a área foliar de seis espécies de leguminosas forrageiras, individualmente, com base em imagens digitais analisadas com o pacote LeafArea, software R. No experimento, avaliou-se 100 folhas aleatoriamente retiradas das seguintes espécies: *C. juncea*, *C. ensiformis*, *C. cajan*, *D. lablab*, *M. cinereum* e *M. aterrima*, sendo utilizadas 70 folhas para ajuste dos modelos e o restante (30) para validação do modelo, tendo como variáveis explicativas o comprimento e a largura do folíolo central. Imagens digitais de cada folha foram processadas para estimar a área foliar. Essas estimativas foram utilizadas para ajustar os modelos, linear e não linear, de predição da área foliar. Para as seis espécies, o modelo linear múltiplo de segundo grau completo pode ser usado para predizer a área foliar em função do comprimento e da largura do folíolo central, apresentando  $R^2 = 0,98$  e erro médio absoluto inferior a 9%. Nos modelos, o efeito da largura do folíolo é, em geral, maior que o comprimento. O pacote R LeafArea mostrou ser uma ferramenta muito eficiente na estimativa da área foliar através da execução do software ImageJ®, com alta precisão e fácil calibração.

**Palavras-chave:** *Crotalária*, *Fabaceae*, *Imagem digital*.

<sup>1</sup> Instituto Federal Goiano – Campus Urutaí. E-mail: hygor\_amaral@hotmail.com

<sup>2</sup> Agradecimento ao IF Goiano pelo apoio financeiro.



## **Novo método estocástico local para predição espacial de dados fitopatogênicos**

**Anderson Rodrigo da Silva<sup>1 2</sup>**

**Ana Paula Alencastro Silva<sup>1 2</sup>**

**Lauro Joaquim Tiago Neto<sup>1 2</sup>**

Dados fitopatogênicos espaciais, como densidade populacional de nematóides, nem sempre apresentam estacionariedade, isotropia e distribuição gaussiana, exigindo técnicas e modelos complexos de krigagem. Alguns métodos determinísticos, como o inverso do quadrado da distância, também podem ser aplicados para prever a infestação, mas o resultado é limitado a soluções gráficas (mapeamento). O objetivo desse trabalho foi adaptar um método computacional de predição, chamado janela circular móvel de raio variável, que baseia-se em outros dois: krigagem da janela móvel e inverso do quadrado da distância. O algoritmo foi desenvolvido para atender a uma função objetivo que minimiza o índice de variação das predições. Para as análises foi desenvolvido um código em R. A saída é mapeada para fornecer uma ferramenta útil para a análise exploratória espacial, incluindo o alcance de dependência espacial através do raio, que é calculado em cada local alvo, bem como o erro padrão dos valores preditos. O método não faz quaisquer suposições sobre o processo espacial, sendo uma alternativa para lidar com a heterogeneidade espacial.

**Keywords:** *Moving window kriging, Software R, Predição espacial.*

---

<sup>1</sup> Instituto Federal Goiano – Campus Urutaí. e-mail: anderson.silva@ifgoiano.edu.br

<sup>2</sup> Agradecimento ao IF Goiano e CNPq (processo 428089/2016-0) pelo apoio financeiro.

## Nonlinear growth models: an alternative to ANOVA in tomato trials evaluation

Alessandro Dal'Col Lúcio<sup>1 3</sup>

Bruno Giacomini Sari<sup>2 3</sup>

Tiago Olivoto<sup>2 3</sup>

Dionatan Ketzer Kryszun<sup>2 3</sup>

Maria Inês Diel<sup>2 3</sup>

The aim was to use growth models as an alternative analysis of experiments with tomatoes. The data were obtained from experiments of tomatoes carried out in two growing seasons. In each harvest the number and mass of fruits per plant were assessed and the Brody, Gompertz, Logistic and von Bertalanffy growth models were fitted as a function of the accumulated values per plant in each harvest and days after the transplant of the seedlings. The assumptions of normality, homogeneity and independence of residuals were verified by the Shapiro-Wilk, Breusch-Pagan and Durbin-Watson tests, respectively. The intrinsic and parametric nonlinearity was calculated according to Bates and Watts methodology. Among the models tested, the one that presented a high  $R^2$  and less non-linearity was selected. The parameters of the selected model were compared by the F test. From the second, third and fourth order derivatives with respect to days after transplant, the maximum acceleration point, inflection point, maximum deceleration point, and point of asymptotic deceleration were estimated. All assumptions regarding residual distributions were met. Among the tested models, the logistic model had the best goodness-of-fit for both variables. Therefore, it was possible to test hypotheses to compare the parameters by the F test, increasing the inferential value of the results. The critical points allowed assessing of the productive precocity, crop concentration period and productive behavior of the genotypes during the productive period. These results present that growth models may be an alternative to analysis of variance in experiments with horticultural crops.

**Keywords:** *Solanum lycopersicum*; Nonlinear regression; Productive precocity; Productive behavior; Logistic model.

<sup>1</sup> Departamento de Fitotecnia/CCR – UFSM. e-mail: adlucio@ufsm.br

<sup>2</sup> Programa de Pós-Graduação em Agronomia - UFSM.

<sup>3</sup> Agradecimento ao CNPq pelo apoio financeiro.

## Número de plantas para o diagnóstico de multicolinearidade em caracteres produtivos de centeio

Ismael Mario Marcio Neu<sup>1</sup>

Alberto Cargnelutti Filho<sup>2</sup>

Gabriela Görgen Chaves<sup>3</sup>

Cirineu Tolfo Bandeira<sup>4</sup>

Jéssica Andiar Kleinpaul<sup>5</sup>

O objetivo deste trabalho foi determinar o número de plantas necessárias para estimar o número de condição (NC) da matriz de correlação entre caracteres produtivos, em duas cultivares de centeio (BRS Progresso e Temprano). Em ensaios de uniformidade, conduzidos em campo, foram coletadas, aleatoriamente, 100 plantas por cultivar. Em cada planta foram mensurados os caracteres produtivos: massas fresca e seca de parte aérea sem espiga, número de espiguetas, massa fresca de espigas, número de grãos e produtividade de grãos. Para cada cultivar, com base na amostra mestre (n=100 plantas) foi calculado a matriz de correlação e determinado o NC. Foram planejados 86 tamanhos de amostra. O primeiro tamanho planejado foi de 15 plantas e os demais foram obtidos com o incremento de uma planta, até o último tamanho contendo 100 plantas. Para cada tamanho de amostra foram realizadas 2.000 reamostragens, com reposição. Para cada reamostra foi determinado o NC. Assim, obteve-se para cada tamanho de amostra, 2.000 estimativas de NC. Determinou-se, para cada tamanho de amostra, o percentil 97,5, a média e o percentil 2,5, sendo essas estatísticas plotadas em gráficos. Há variabilidade no grau de multicolinearidade entre as matrizes de correlação da amostra mestre (n=100 plantas) para as duas cultivares de centeio, com NC de 2303,48 e 693,46 para BRS Progresso e Temprano, respectivamente. Em ambas as cultivares, a média das 2.000 estimativas de NC decresceu à medida que o tamanho de amostra planejado aumentou, com tendência de estabilização a partir de tamanhos de amostra superiores a 50 plantas.

**Palavras-chave:** *Intervalo de confiança, Reamostragem, Secale cereale.*

<sup>1</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>2</sup> Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com. Bolsista de Produtividade em Pesquisa 1C-CNPq - Processo: 305057/2013-8

<sup>3</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>4</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>5</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CNPq. Agradecimentos: Ao CNPq e CAPES, pelas bolsas concedidas.

## **Obtenção de Grupos de Locais para Experimentos com Cana-de-Açúcar: uma comparação do Teste de Scott-Knott e da Análise de Agrupamentos**

**Cristiane Mariana Rodrigues da Silva**<sup>1</sup>

**Carlos Tadeu dos Santos Dias**<sup>2</sup>

No ano de 2004 foram instalados ensaios em 11 Unidades Experimentais pertencentes ao Centro de Tecnologia Canavieira, dentro do estado de São Paulo, e houve a necessidade de se saber se é possível a redução deste número. Sendo assim, este trabalho teve por objetivo, aplicar as técnicas da Análise de Agrupamento ("Cluster Analysis") e o teste de Scott-Knott na identificação da existência de grupos de Unidades Industriais, visando à diminuição do número de experimentos e, por conseguinte, visando ao menor custo operacional. Os métodos de comparação múltipla baseados em análise de agrupamento univariada, têm por objetivo separar as médias de tratamentos que, para esse estudo foram médias de locais. Neste caso, o teste de Scott-Knott, utilizando a variável TPH5C (tonelada de POL por hectare estimada para 5 cortes), por se tratar de uma variável calculada a partir de outras variáveis como POL (porcentagem de açúcar), TCH (tonelada de cana por hectare) e FIB (porcentagem de fibra). Na Análise de Agrupamentos, por meio do Método de Ligação das Médias, mostrou-se mais confiável porque a análise envolvia simultaneamente as três variáveis que se seguem: TCH, POL, e FIB. Comparando-se o teste de Scott-Knott com a Análise de Agrupamentos, confirmaram-se os agrupamentos entre os locais L020 (Ribeirão Preto) e L076 (Assis) e os locais L045 (Ribeirão Preto) e L006 (Jau). Concluiu-se, portanto, que podem ser eliminadas dos experimentos duas unidades experimentais, optando por L020 ou L076, e L045 ou L006, ficando essa escolha, a critério do pesquisador, podendo assim, reduzir seu custo operacional.

**Palavras-chave:** *Análise de Agrupamentos, Cana-de-Açúcar, Scott-Knott.*

<sup>1</sup> PPG em Estatística e Experimentação Agronômica, ESALQ/USP, cmrsilva@usp.br.

<sup>2</sup> Prof. Titular, Dep. Ciências Exatas, ESALQ/USP, ctsdias@usp.br.

---

## **Padrões e métodos para análise da distribuição espacial de larva-minadora em cultivo de alface hidropônico**

**Wilhan Valasco dos Santos<sup>1 2</sup>**

**Brunna Rithielly Rezende<sup>1 2</sup>**

**Hygor Amaral Santanat<sup>1 2</sup>**

**Anderson Rodrigo da Silva<sup>1 2</sup>**

A detecção e predição da distribuição espacial de pragas de plantas cultivadas pode ser uma ferramenta facilitadora do manejo, indicando locais no campo que mais necessitam de intervenções com pesticidas. O presente estudo teve como objetivo o estudo da distribuição da larva-minadora na cultura da alface hidropônico e comparar a precisão de diferentes métodos de análise de distribuição espacial. O trabalho em questão foi realizado em julho de 2017 na estação experimental do IPA. As mudas preparadas em espuma fenólica foram cultivadas em quatro bancadas de 6 x 1,5 m, com diferentes soluções nutritivas comerciais. Dados de infestação de larva-minadora (número de larvas por folha) e de componentes de produção (número de folhas, peso de folhas, diâmetro e comprimento de caule) foram coletados em um grid amostral de 0,25 x 0,5 m, em cada bancada. Foram adotados os seguintes métodos de análise espacial de dados: geoestatística clássica (krigagem ordinária), correlograma de Mantel, índice de agregação de Morisita, interpolação pelo inverso do quadrado da distância, método da janela circular móvel de raio variável. Todos os métodos de análise de distribuição espacial foram convergentes quanto ao padrão de distribuição de larva-minadora em alface hidropônico: padrão regular. As soluções nutritivas distintas não afetam o padrão de distribuição da larva-minadora.

**Palavras-chave:** *Autocorrelação espacial, Hidroponia, Liriomyza huidobrensis.*

---

<sup>1</sup> Instituto Federal Goiano – Campus Urutaí. E-mail: hygor\_amaral@hotmail.com

<sup>2</sup> Agradecimento ao IF Goiano pelo apoio financeiro.

## Parametrização do modelo Stanford & Smith na descrição da mineralização do carbono da palha de aveia

Edilson Marcelino Silva<sup>1 3</sup>

Thais Destefani Ribeiro<sup>2 3</sup>

Joel Augusto Muniz<sup>2 3</sup>

O modelo não linear Stanford & Smith, dado por  $C_t = C_0(1 - \exp(-kt))$ , é utilizado em estudos de mineralização do carbono, em que  $C_t$  é a porcentagem do carbono mineralizado até o tempo  $t$  em dias,  $C_0$  a quantidade de carbono potencialmente mineralizável e  $k$  a taxa de mineralização. Uma parametrização interessante neste modelo é considerar o parâmetro  $k$  como  $\ln(2)/v$ , em que  $v$  é o tempo de meia vida do carbono potencialmente mineralizável. O objetivo deste trabalho foi avaliar a não linearidade do modelo na forma original e com a parametrização de acordo com as medidas propostas por Bates e Watts (1980), na descrição do  $\text{CO}_2$  mineralizado em palha de aveia na superfície do solo. As análises foram realizadas no software R. Os resultados indicaram que a parametrização reduziu o desvio da linearidade, e portanto, esta forma deve ser indicada em estudos de mineralização de  $\text{CO}_2$  tendo em vista a interpretação biológica do parâmetro  $v$ . Para a palha de aveia na superfície do solo, o tempo de meia vida do carbono potencialmente mineralizável foi estimado em 24 dias, indicando que neste tempo 50% da quantidade de  $\text{CO}_2$  potencialmente mineralizável se encontra mineralizada.

**Palavras-chave:** *não linearidade, interpretação biológica, tempo de meia vida.*

---

<sup>1</sup> DES – UFLA. e-mail edilsonmg3@hotmail.com

<sup>2</sup> DES – UFLA

<sup>3</sup> Agradecimento a CAPES e Fapemig pelo apoio financeiro.

# Agricultura de precisão: uma aplicação dos GAMLSS

Ana Julia Righetto <sup>1</sup>

Thiago Gentil Ramires <sup>2</sup>

Luiz Ricardo Nakamura <sup>3</sup>

Pedro Lian D. B. Castanho <sup>4</sup>

Taciana Villela Savian <sup>5</sup>

O cultivo de cana-de-açúcar vem se concentrando em vários países devido à sua diversidade de uso. A modernização da agricultura permitiu alta produtividade, que é afetada pela invasão de ervas daninhas. Com a agricultura sustentável, o uso de herbicidas tem sido cada vez mais evitado na sociedade, exigindo métodos de controle mais eficazes. Neste trabalho, propomos um modelo estatístico capaz de identificar a invasão de ervas daninhas no campo, usando quatro espectros de cores como variáveis regressoras ue foram obtidas por uma câmera multiespectral montada em um veículo aéreo não tripulado. Com a identificação exata da infestação, é possível realizar a gestão no campo com aplicações de herbicidas nos lugares exatos, evitando assim o aumento do custo de produção e até mesmo dispensando o uso de herbicidas, efetuando a remoção mecânica das ervas. Resultados mostram que no campo experimental foi possível reduzir em 57% o uso de herbicida.

**Palavras-chave:** *Regressão logística multinomial, Agricultura moderna, Erva daninha.*

---

<sup>1</sup>IAPAR. e-mail: [ajrighetto@gmail.com](mailto:ajrighetto@gmail.com)

<sup>2</sup>DAMAT - UTFPR.

<sup>3</sup>INE - UFSC.

<sup>4</sup>RAÍZEN.

<sup>5</sup>LCE - USP/ESALQ..

## Regressão quantílica não linear na descrição de crescimento de rã touro na fase de recria.

Kelly Pereira de Lima <sup>1 3</sup>

Charles Shalimar Felipe da Silva <sup>2 3</sup>

Jaqueline Gonçalves Fernandes <sup>2 3</sup>

Marcela de Marillac Carvalho <sup>2 3</sup>

Augusto Ramalho de Moraes <sup>2 3</sup>

Análises de crescimento explicam o crescimento animal ao longo do tempo, gerando informações referentes ao comportamento do crescimento animal. Os modelos não lineares podem ser apropriados descrever estas curvas, pois seus parâmetros apresentam interpretação biológica. Contudo, estes modelos proporcionam informações em função da média, e estão subordinados a problemas no ajuste por possíveis valores atípicos ou assimetria na distribuição dos dados. Desta maneira, a regressão quantílica pode contornar estes problemas, e ainda permite estimativas de diferentes quantis, gerando resultados mais completos e robustos. Assim, objetivo desse trabalho é o ajuste de modelos de regressão quantílica não linear para descrever o crescimento em rã-touro na fase de recria. Além disso, comparar os resultados com os obtidos pelo modelo não linear. Os dados usados são provenientes de um experimento Rodrigues (2007) com curva de crescimento em peso e idade de rã-touro na fase de recria. Foram utilizados 300 rãs-touro que foram pesadas durante 294 dias em intervalos de 14 dias. Para a regressão quantílica usou-se os quantis 0,25 ; 0,50 e 0,75 para o modelo gompertz. Conclui-se que foi possível utilizar várias curvas para obter uma visão mais completa da relação entre as variáveis estudadas. Além disso, verificou-se que as estimativas obtidas pelas regressão quantílica não linear foi que o peso máximo é de 308,51, sendo maior no quantil 0,75 (328,85) e menor no quantil 0,25 (297,94).

**Palavras-chave:** *Crescimento, Quantis, Software R.*

---

<sup>1</sup>DES -UFLA. e-mail: *kelly.lima.88@gmail.com*

<sup>2</sup>DES - UFLA.

<sup>2</sup>DES - UFLA.

<sup>2</sup>DES - UFLA.

<sup>2</sup>DES - UFLA.

<sup>3</sup>Agradecimento ao CNPq, Capes e FAPMIG pelo apoio financeiro.



## **Relações lineares em caracteres de cultivares de milho de ciclo precoce**

**Eder de Souza<sup>1</sup>**

**Alberto Cargnelutti Filho<sup>2</sup>**

**Rosana Marzari Thomasi<sup>3</sup>**

**Daniela Lixinski Silveira<sup>4</sup>**

**Guilherme Padilha<sup>5</sup>**

O objetivo deste estudo foi avaliar, em cultivares de milho de ciclo precoce, a relação entre a produtividade de grãos e os caracteres agronômicos. No ano agrícola 2016/2017 foi conduzido um experimento com treze cultivares de milho (*Zea mays* L.) de ciclo precoce. O delineamento experimental utilizado foi blocos ao acaso, com três repetições. Em cada unidade experimental foram mensurados os caracteres, altura de planta na colheita, altura de espiga na colheita, posição relativa da espiga, número de plantas, número de espigas, índice de espigas e a produtividade de grãos, a 13% de umidade. Os dados foram submetidos à análise de variância, e realizados os testes de Lilliefors e Bartlett para verificar o atendimento, respectivamente, da normalidade dos erros e da homogeneidade das variâncias residuais. Realizou-se o teste de comparação de médias de Scott Knott, a 5% de probabilidade, para comparar as médias das cultivares. Foi determinada a matriz de correlação fenotípica e verificada a multicolinearidade. Foi realizada a análise de trilha, considerando a produtividade de grãos como caractere principal. Os caracteres número de espiga e altura de espiga apresentaram efeito direto positivo sobre a produtividade de grãos e podem ser utilizados na seleção indireta de plantas.

**Palavras-chave:** *Análise de trilha, Seleção indireta, Zea mays* L.

<sup>1</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PIBIC/CNPq/UFSM.

<sup>2</sup> Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com. Bolsista de Produtividade em Pesquisa 1C-CNPq - Processo: 305057/2013-8

<sup>3</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PIBIC/CNPq/UFSM.

<sup>4</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>5</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PIBIC/CNPq/UFSM.

Agradecimentos: Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelas bolsas concedidas.

## **Relações lineares em caracteres de cultivares de milho de ciclo superprecoce**

**Guilherme Padilha<sup>1</sup>**  
**Alberto Cargnelutti Filho<sup>2</sup>**  
**Daniela Lixinski Silveira<sup>3</sup>**  
**Rosana Marzari Thomasi<sup>4</sup>**  
**Eder de Souza<sup>5</sup>**

O objetivo deste trabalho foi avaliar, em genótipos de milho de ciclo superprecoce, a relação entre os caracteres morfológicos e produtivos e a produtividade de grãos. No ano agrícola 2016/2017 foi conduzido um experimento com dez cultivares de milho (*Zea mays* L.) de ciclo superprecoce, no delineamento de blocos ao acaso, com três repetições. Foram mensurados os caracteres morfológicos (altura de planta na colheita, altura de espiga na colheita, posição relativa da espiga) e produtivos (número de plantas, número de espigas, índice de espigas) e a produtividade de grãos, a 13% de umidade. Os dados foram submetidos à análise de variância. Foram realizados os testes de Lilliefors e Bartlett para verificar o atendimento dos pressupostos de normalidade dos erros e de homogeneidade das variâncias residuais, respectivamente. A acurácia seletiva foi utilizada para avaliação da precisão experimental. As médias das cultivares foram agrupadas pelo teste de Scott Knott a 5% de probabilidade. Foi determinada a matriz de correlação fenotípica e realizado o diagnóstico de multicolinearidade. A análise de trilha foi realizada, considerando a produtividade de grãos como caractere principal e os demais, como caracteres explicativos. Em cultivares de milho de ciclo superprecoce, os caracteres altura de espiga e número de espigas apresentam correlação direta positiva sobre a produtividade de grãos e podem ser utilizados para a seleção indireta de plantas em programas de melhoramento genético.

**Palavras-chave:** *Correlação fenotípica, Multicolinearidade, Seleção indireta, Zea mays L.*

<sup>1</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PIBIC/CNPq/UFSM.

<sup>2</sup> Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com. Bolsista de Produtividade em Pesquisa 1C-CNPq - Processo: 305057/2013-8

<sup>3</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>4</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PIBIC/CNPq/UFSM.

<sup>5</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PIBIC/CNPq/UFSM.

Agradecimentos: Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelas bolsas concedidas.

## **Seleção direta e indireta de caracteres relacionados a resistência ao nematoide das galhas em progênies de alface**

**Renato Domiciano Silva Rosado<sup>1 3</sup>**

**Sylmara Fernando de Pol Mayer<sup>2 3</sup>**

**Luiz Antônio Augusto Gomes<sup>2 3</sup>**

**Deborah Abreu Queiroz<sup>2 3</sup>**

**Daniele de Fátima Oliveira<sup>2 3</sup>**

Nos programas de melhoramento de alface a resistência ao nematoide das galhas para sistemas orgânicos é um desafio. Com o auxílio de parâmetros genéticos, como a herdabilidade e o ganho com a seleção, pode-se identificar genótipos superiores em gerações precoces. Assim, o objetivo desse trabalho foi predizer ganhos por seleção e identificar progênies superiores de alface F2:4, oriundas do cruzamento entre as cultivares Salinas 88 e Colorado, quanto à homozigose para a resistência ao nematoide das galhas. O delineamento utilizado foi o de blocos casualizados com quatro repetições, em que foram avaliadas 27 progênies, com três testemunhas, os genitores e a cultivar Regina 71. Foram avaliadas as variáveis: número de ovos (NO), índice de reprodução (IR), fator de reprodução (FR), incidência de galhas (IG). Foram identificadas variabilidade genética em todas as variáveis. Herdabilidades maiores que 75% foram observadas em NO, IR e FR. As progênies AFX 024D 1172 3151, AFX 024D 1181 3182, AFX 024D 1211 3167, AFX 024D 1241 3277 e AFX 024D 1309 3487 devem ser selecionados para avançarem no programa de melhoramento visando resistência ao nematoide das galhas. A seleção com base número de ovos, índice de reprodução e fator de reprodução, pode promover maiores ganhos de seleção em gerações mais avançadas. Se o interesse for pelo método não destrutivo a seleção das progênies por seleção direta em IG deve ser levada em consideração, já que, os ganhos indiretos são na mesma direção e superiores ao IG.

**Palavras-chave:** *Índices de seleção, melhoramento de alface, resistência a Meloidogyne.*

---

<sup>1</sup> Departamento de Estatística Aplicada e Biometria – UFV. e-mail [renato.rosado@ufv.br](mailto:renato.rosado@ufv.br)

<sup>2</sup> Departamento de Agricultura Geral – UFLA.

<sup>3</sup> Agradecimento a CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## Spatial variability of potentially toxic elements Ba, Ni and Zn in tropical soils

Gener Tadeu Pereira<sup>a</sup>, Livia Arantes Camargo<sup>a</sup>, José Marques Júnior<sup>a</sup>, Luís Reynaldo Ferracciú Alleoni<sup>b</sup>

<sup>a</sup>FCAV/UNESP. São Paulo State University, Jaboticabal SP, Brazil.  
(genertp@fcav.unesp.br)

<sup>b</sup>College of Agriculture Luiz de Queiroz (ESALQ), University of Sao Paulo (USP), Piracicaba SP, Brazil.

Indicator kriging is a non-parametric type of conditional estimation, which uses the position and values of the data to produce a local distribution rather than a global distribution of the analyzed property. One example could be performed into the probability of occurrence of potentially toxic elements (EPTs) Ba, Ni and Zn. For that, soil samples were collected from three geomorphic surfaces and analyzed for chemical, physical, and mineralogical properties. Geostatistical analysis was used to characterize the spatial variability of EPTs. Indicator kriging was used for the preparation of probabilistic maps and cut levels were chosen to delimit spatial distribution patterns. The cut-off values were defined according to the guideline values of each EPT (values according to the Brazilian legislation). The variable Ba presented a higher probability of occurrence of higher levels than the VRQ in part of SG II and SG III.. The Ni and Zn variables were more likely to occur higher than the VRQ in part of SG II and SG III and a higher probability of occurrence of levels higher than SG in SG III. These results show that the boundary region between SG II and SG III and SG III already present higher levels than those considered natural to the soil and require more attention regarding the prevention of elevation of the EPT levels studied at intervention levels. It is concluded that the indicative kriging technique was efficient in the definition of regions with different levels of risk of contamination by the EPTs studied.

---

## MODELAGEM DO CRESCIMENTO DE POEDEIRAS COMERCIAIS, CORPO E DE SEUS COMPONENTES, A PARTIR DE OBSERVAÇÕES "in vivo"

Malheiros, E. B.

Alves Junior, W.

Sakomura, N. K.

Alves, E. P

O modelo de crescimento de aves de Emmans (1981) descreve o animal por métodos destrutivos. O DXA (absorciometria de raios-X) é um método alternativo que permite, a partir da composição corporal *in vivo*, estudar variações do crescimento animal (SCHOLZ et al., 2015). Ensaio realizado: 90 aves - Dekalb Brown, avaliações: 1-126 dias.

Elaborou-se modelos de predição da composição corporal *in vivo*, a partir de das variáveis obtidas no DXA: X1:MT - Massa total, X2:MG - Massa gorda, X3:MM - massa magra (água + proteínas), X4:CMO - conteúdo mineral ósseo e, para avaliar quais delas interferem na composição corporal do animal (Y), usou o modelo:  $\ln(Y) = B_0 + B_1 \ln(X_1) + B_2 \ln(X_2) + B_3 \ln(X_3) + B_4 \ln(X_4) + \epsilon$ ; Y - cada variável estudada "in vivo": PCLP - peso do corpo livre de penas, MM - matéria mineral, GD - massa de gordura, PB - teor de proteína bruta e PACLP - peso de água do corpo livre de penas. Pelo método "stepwise" selecionou-se os modelos: PCLP = f(MT); MM = f(CMO, MM); GD = f(CMO, MG, MM, MT); PB = f(MM) e PACLP = f(MM, MT).

Com os valores estimados da composição corporal, ajustou-se o modelo para descrever o crescimento das aves, Gompertz,  $W_t = W_m \times \exp(-\exp(\ln(-\ln(W_t / W_m)) - (B \times t)))$ , onde:  $W_t$ =peso-inicial,  $W_m$ =peso-maturidade, B = taxa crescimento,  $t^*$  - ponto de inflexão.

**Palavras-chave:** Modelagem de crescimento, poedeiras comerciais; método não destrutivo.

## Tamanho de parcela para avaliar a massa de matéria fresca de aveia preta e ervilhaca

Andréia Procedi<sup>1</sup>

Alberto Cargnelutti Filho<sup>2</sup>

Rafael Vieira Pezzini<sup>3</sup>

Jéssica Maronez de Souza<sup>4</sup>

Aline Sielo Holkem<sup>5</sup>

Dimensionar adequadamente o tamanho de parcela, a ser utilizado no experimento, é fundamental para obtenção de resultados fidedignos dos tratamentos em avaliação. O objetivo deste trabalho foi determinar o tamanho ótimo de parcela ( $X_o$ ) para avaliar a massa de matéria fresca de aveia preta (*Avena strigosa* Schreb) e de ervilhaca (*Vicia sativa* L.), em cultivos isolados e consorciados. Em cada uma das composições de densidades de semeadura de aveia preta (AP) e de ervilhaca (ER) (C1 = 100% de AP; C2 = 75% de AP + 25% de ER; C3 = 50% de AP + 50% de ER; C4 = 25% de AP + 75% de ER; e C5 = 100% de ER), foram conduzidos três ensaios de uniformidade de dimensão 6m×6m, totalizando 15 ensaios. Cada ensaio foi dividido em 36 unidades experimentais básicas (UEB) de 1m×1m, formando uma matriz de seis linhas e seis colunas. Em cada UEB foram cortadas as plantas, junto à superfície do solo, e pesada a massa de matéria fresca. Para cada ensaio de uniformidade foi determinado o coeficiente de autocorrelação espacial de primeira ordem, a variância, a média e o coeficiente de variação. O  $X_o$  foi determinado por meio do método da curvatura máxima do modelo do coeficiente de variação. Há variabilidade do  $X_o$  entre as composições de densidades de plantas de aveia preta e ervilhaca. O  $X_o$  oscilou entre 2,92 m<sup>2</sup> para a composição de 25% de AP + 75% de ER e 4,01 m<sup>2</sup> para a composição de 100% de ER.

**Palavras-chave:** *Avena strigosa*, *Ensaio de uniformidade*, *Planejamento experimental*, *Vicia sativa*.

<sup>1</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista PROBIC/FAPERGS/UFSM.

<sup>2</sup> Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com. Bolsista de Produtividade em Pesquisa 1C-CNPq - Processo: 305057/2013-8

<sup>3</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>4</sup> Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista CAPES.

<sup>5</sup> Curso de Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Bolsista IC/CNPq - Projeto: 401045/2016-1 - Processo: 119164/2017-5

Agradecimentos: Ao CNPq, CAPES e FAPERGS, pelas bolsas concedidas.

# Uma abordagem às equações diferenciais a partir do estudo do modelo não linear Chanter

Pollyane Vieira da Silva <sup>1</sup>

Taciana Villela Savian <sup>2</sup>

Muitos problemas importantes e significativos da engenharia, das ciências físicas e das ciências sociais, formulados em termos matemáticos, exigem a determinação de uma função que obedece a uma equação que contém uma ou mais derivadas da função desconhecida. Estas equações são chamadas de equações diferenciais.

O estudo de equações diferenciais iniciou-se do cálculo, com Isaac Newton (1642-1727) e Gottfried Wilhelm Leibniz (1646-1716) no século XVII (BOYCE; DIPRIMA, 1992). Atualmente há vários estudos sobre o crescimento de frutos, como Ribeiro et al (2018) estudou frutos de pereira, Fernandes et al (2017) estudou frutos do cafeeiro, esse crescimento é descrito por curvas obtidas dos modelos de regressão não lineares como modelo Logístico, modelo Gompertz, entre outros. As equações desses modelos não lineares de crescimento são soluções de equações diferenciais, cujos parâmetros podem ser interpretados biologicamente.

O modelo Chanter é um modelo de crescimento proposto por Dennis Osborne Chanter em 1976. É um híbrido entre os modelos Logístico e o Gompertz cujos parâmetros podem ser interpretados similarmente (FRANCE; THORNLEY, 1984). O adjetivo híbrido dado ao modelo Chanter provém da relação entre a equação diferencial do modelo Logístico e do modelo Gompertz.

Neste trabalho o objetivo é apresentar a solução da equação diferencial do modelo Chanter e uma aplicação do modelo Chanter na descrição do crescimento do fruto do cacaueiro.

**Palavras-chave:** *Modelo Chanter, Equação diferencial, Cacaueiro.*

---

<sup>1</sup>LCE - ESALQ. e-mail: *vieira\_polly@usp.br*

<sup>2</sup>LCE - ESALQ. e-mail: *tsavian@usp.br*

# Uma apresentação didática das Bayesian Networks no software R

Tatiane Carvalho Alvarenga <sup>1 3</sup>

Renato Ribeiro de Lima <sup>2 3</sup>

As *Bayesian Networks* (BN) foram propostas na década de 80 por Judea Pearl, então conhecido por defender o conhecimento probabilístico no campo da inteligência artificial. As aplicações dessa metodologia tem se estendido nas diferentes áreas do conhecimento, sendo indicada para lidar com problemas que tratam incertezas. As *Bayesian Networks* são modelos gráficos (*graphical models*), os quais consistem na representação gráfica (grafo) e probabilística (distribuições de probabilidade condicionais e conjunta) das variáveis. O objetivo deste trabalho é apresentar de uma forma didática essa metodologia, suas implementações (algoritmos de *machine learning*) e funcionalidades utilizando o software R. Os resultados das aplicações de BN têm-se mostrado eficiente, o que justifica sua superioridade em capturar (e expressar) relações mais complexas entre as covariáveis e a variável resposta, especialmente no cenário atual de big data.

**Palavras-chave:** *Inteligência Artificial. Machine learning. Modelos Gráficos.*

---

<sup>1</sup>DES - UFLA. e-mail: [tatianedex@posgrad.ufla.br](mailto:tatianedex@posgrad.ufla.br)

<sup>2</sup>DES - UFLA.

<sup>3</sup>Agradecimentos a CAPES e FAPEMIG pelo apoio financeiro.



## Uso de Árvores Aleatórias para Classificação Sensorial de Arroz Cozido

Rafael Lima de Moraes <sup>1</sup>

Dr. George Freitas von Borries <sup>2</sup>

Dra. Priscila Zaczuk Bassinello <sup>3</sup>

Ma. Rosangela Nunes Carvalho <sup>3</sup>

A predição da classificação de textura do arroz cozido possui interesse comercial direto e vem sendo estudada pela Embrapa Arroz e Feijão (CNPAG-Go). O presente estudo foca em modelar a classificação de pegajosidade por meio de medidas de teor de amilose e viscosidade do grão de arroz cozido utilizando um classificador por floresta aleatória (*Random Forest*) e com base na agregação de múltiplas árvores de decisão. Esta técnica é muito conhecida no cenário de aprendizado de máquina e se destaca por dispensar grandes suposições aos dados.

Através do uso de componentes principais, a dimensionalidade do estudo foi reduzida para aplicação de floresta aleatória na combinação de diversas árvores para construção de um modelo de classificação da variável resposta que indica o nível de pegajosidade do arroz. O conjunto de dados *out-of-bag* (OOB), gerado do processo de construção da floresta serve de cálculo do erro de predição populacional dos dados.

Neste estudo foram detalhados os resultados da classificação sensorial de pegajosidade para o arroz cozido de terras altas colhido no ano de 2014. Para esse conjunto o modelo apresentou erro OOB de 15,28%, o que corresponde a 11 observações erroneamente classificadas em um total de 72. A floresta aleatória mostrou-se capaz de captar estruturas não lineares das categorias. As observações com erros de classificação OOB indicam a limitação do modelo em generalizar regiões de confundimento.

**Palavras-chave:** *Árvores de Decisão, Erro de classificação, Floresta Aleatória, Random Forest, Visualização de dados.*

---

<sup>1</sup>Mestrando, Departamento de Administração, Universidade de Brasília. E-mail: [rafaelestatistico@gmail.com](mailto:rafaelestatistico@gmail.com)

<sup>2</sup>Professor, Departamento de Estatística, Universidade de Brasília, e-mail: [gborries@unb.br](mailto:gborries@unb.br)

<sup>3</sup>Pesquisadora, EMBRAPA Arroz e Feijão.

<sup>3</sup>Analista, EMBRAPA Arroz e Feijão.

## COMPARAÇÃO EM ANÁLISE DE AGRUPAMENTO UM ESTUDO DE CASO

Mácio Augusto de Albuquerque <sup>(1)</sup>

Kleber Napoleão Nunes de Oliveira Barros <sup>(1)</sup>

Hiago José Andrade de Albuquerque Martins <sup>(1)</sup>

Pedro Augusto Ferreira Silva <sup>(1)</sup>

**Resumo** - Objetivou-se propor uma sistemática para o estudo e a interpretação da comparação dos métodos em análise de agrupamento, através de vários algoritmos de agrupamento em dados de vegetação. Utilizou-se dados provenientes de um levantamento na Mata da Silvicultura, da Universidade Federal de Viçosa-MG. Para análise de agrupamento foram estimadas as matrizes de distância de *Mahalanobis* com base nos dados originais e via reamostragem “*bootstrap*” e aplicados os métodos da ligação simples e ligação completa. Para a detecção de associação entre os métodos foi aplicado o teste *qui-quadrado* ( $\chi^2$ ) a 1% e 5% de probabilidade. Observou-se que os métodos considerados são coincidentes, com base nas matrizes de *Mahalanobis* a partir dos dados originais e “*bootstrap*”. A sistemática proposta é promissora para o estudo e a interpretação da estabilidade dos métodos de análise de agrupamento.

**PALAVRAS-CHAVE:** análise multivariada, bootstrap, métodos hierárquicos.

---

<sup>(1)</sup> DEST – UEPB. e-mail macioaa@uol.com.br.

# Bioinformática

## **Controlador Fuzzy híbrido aplicado em genótipos de arroz: Análise de adaptabilidade e estabilidade**

**Márcia Elaine Guimarães Lana**<sup>1 4</sup>

**Weverton Gomes da Costa**<sup>2 4</sup>

**Alexandre Gomes Ferraz**<sup>2 4</sup>

**Gustavo Henrique Pataro de Oliveira**<sup>3 4</sup>

**Plínio César Soares**<sup>3 4</sup>

Diversas metodologias de adaptabilidade e estabilidade estão propostas na literatura. No entanto, muitas delas possuem interpretações complexas, principalmente aquelas que utilizam grande quantidade de parâmetros. Assim, as análises baseadas em inteligência computacional podem auxiliar na interpretação dessas análises. O objetivo desse trabalho foi avaliar o comportamento de genótipos de arroz em diferentes ambientes utilizando um controlador fuzzy híbrido de duas metodologias de adaptabilidade e estabilidade. Foram avaliadas 21 linhagens de arroz do programa de melhoramento genético da Epamig quanto à produtividade de grãos em três localidades, nos anos agrícolas 2012/2013, 2013/2014 e 2014/2015. Foram realizadas análises de variância conjunta dos dados e de adaptabilidade e estabilidade. A análise de variância conjunta indicou que houve interação genótipo x ambiente significativa, então foi realizada a análise de adaptabilidade e estabilidade pelos métodos de Eberhart e Russell e de Lin e Binns modificado. Os parâmetros obtidos foram submetidos ao controlador fuzzy híbrido. Os genótipos 3, 6, 9, 10, 15 e 18 foram classificados como de ampla adaptabilidade para produtividade de grãos. O genótipo 15 (CNAi 9091) foi o que apresentou maior pertinência entre os genótipos de ampla adaptabilidade, sendo que este genótipo tem sobressaído nos últimos anos agrícolas. Para classificação de adaptabilidade desfavorável podemos destacar os genótipos 1, 19 e 20, e os demais genótipos foram classificados como pouco adaptado. Por considerar um maior número de parâmetros, o controlador fuzzy híbrido é uma alternativa interessante para tomar a decisão na recomendação de cultivares em um programa de melhoramento.

**Palavras-chave:** *Oryza sativa L., Tocher, Variabilidade.*

---

<sup>1</sup> DMA – UFV. e-mail: marcia.guimaraesufv@gmail.com

<sup>2</sup> DBG – UFV.

<sup>3</sup> EPAMIG – UFV

<sup>4</sup> Agradecimento ao CNPq, CAPES e à FAPEMIG pelo apoio financeiro.

## **Desempenho de híbridos por meio de simulação das análises fenotípicas dos pais**

**Weverton Gomes da Costa<sup>1 3</sup>**

**Isabela de Castro Sant'anna<sup>2 3</sup>**

**Francyse Edite de Oliveira Chagas<sup>2 3</sup>**

**Antônio Carlos da Silva Júnior<sup>2 3</sup>**

**Cosme Damião Cruz<sup>2 3</sup>**

A seleção genômica ampla associa informações moleculares e fenotípicas e prediz o desempenho de progênies futuras (valor genético genômico predito) com informações moleculares. O objetivo desse trabalho foi utilizar a seleção genômica ampla para escolha apropriada de genitores e predição do desempenho de híbridos não realizados. Foi realizada a simulação de 200 linhagens endogâmicas recombinantes (RILs), podendo gerar até 19900 híbridos. Foi utilizado o valor fenotípico como critério de escolha dos pais. Posteriormente, executou a simulação fenotípica de cinco características: cercosporiose, teor de óleo, plantas acamadas, altura e rendimento. Para simulação genotípica utilizou-se 1000 marcadores com 10 grupos de ligação e com tamanho variável. Para a população de treinamento foi utilizado apenas o valor fenotípico e a genotipagem de todas as linhagens e dos híbridos formados em um dialelo, em que os pais foram previamente submetidos ao um ensaio de competição e para a população de validação foi utilizada a genotipagem de todos os híbridos. Neste modelo foram escolhidos 45 híbridos levando-se em consideração apenas a característica produtividade de grãos, considerada a mais complexa por envolver maior número de genes e ter menor herdabilidade. Dos híbridos escolhidos teve-se que 82,2% estavam classificados no grupo dos ótimos híbridos quando comparado com o valor genotípico real desses indivíduos. A correlação de Pearson foi alta (0,7591) entre o valor genotípico real e o obtido por meio da seleção genômica ampla. Esses resultados comprovam que o modelo de predição construído foi adequado para prever o desempenho de híbridos que não foram realizados.

**Palavras-chave:** *Seleção genômica ampla, predição de híbridos, correlação.*

---

<sup>1</sup> DBG – UFV. e-mail wevertonufv@gmail.com

<sup>2</sup> DBG – UFV.

<sup>3</sup> Agradecimento ao CNPq e à FAPEMIG pelo apoio financeiro.

---

## Estimação da Estrutura de Dependência do Genoma por MCGLM

Francisco J A Fernandes<sup>1</sup>

Júlia Maria Pavan Soler<sup>2</sup>

A identificação de blocos de dependência dentro do genoma humano tem sido um desafio e alvo de várias iniciativas de pesquisa na área de inferência, muitas delas utilizando marcadores genéticos do tipo SNP (*Single Nucleotide Polymorphisms*), com alta densidade dentro do genoma. Este trabalho fez uso de uma modelagem baseada em processos estocásticos baseados em Campos Markovianos e árvores de contexto para inferir esses blocos de dependência usando dados de SNPs de indivíduos independentes. Entretanto, quando temos dados de indivíduos e seus familiares e, conseqüentemente, uma estrutura de dependência entre as observações, tal procedimento não pode ser usado sem uma transformação adequada nos dados. Para tratar essa limitação, propõe-se a utilização de um modelo linear generalizado com estrutura de covariância multivariada (MCGLM – *Multivariate Covariance Generalized Linear Models*), proposto por Bonat e Jørgensen (2016), com flexibilidade para modelar a estrutura de covariância entre indivíduos bem como entre variáveis (genótipos de SNP). Adicionalmente, é considerado um procedimento de seleção de variáveis baseado em testes de correlação parcial. O modelo MCGLM proposto é comparado com a solução estocástica, para dados independentes.

**Palavras-chave:** *Campos Markovianos, Dados de SNP, Modelos Lineares Generalizados.*

---

<sup>1</sup> IME – USP e-mail: fjafern@ime.usp.br

<sup>2</sup> IME – USP e-mail: pavan@ime.usp.br

## Estudo sobre o genoma do Zika vírus no Brasil por meio da análise de ondaletas

Leila Maria Ferreira<sup>1 4</sup>

Kelly Pereira de Lima<sup>2 4</sup>

Thelma Sáfiadi<sup>3 4</sup>

O Zika Virus é considerado um grande problema de saúde pública, pois é uma doença que ainda não tem um tratamento específico e a sua principal forma de combate tem sido o controle populacional do vetor, por meio de políticas públicas educacionais de prevenção. Neste sentido, a utilização de técnicas estatísticas têm auxiliado na investigação do comportamento da doença, bem como fornecendo importantes informações, como a identificação de locais e populações vulneráveis ao avanço dessa enfermidade. O principal objetivo desse trabalho foi de conduzir uma análise do genoma do Zika vírus no Brasil, utilizando a transformada discreta decimada de ondaletas (DWT), sob duas perspectivas de ondaletas ( Haar e Daubechies) com intuito de verificar o nível de decomposição mais suave, observar o comportamento de cada nível de decomposição e identificar a quantidade de energia em cada nível. De acordo com os resultados, o nível 4 possui uma melhor suavização com relação ao genoma do Zika, ou seja, esse nível possui o maior detalhamento (capacidade de visualizar componentes escondidos no genoma) da série analisada. Além disso, as duas ondaletas analisadas apresentaram comportamento semelhantes na decomposição da série e os valores da energia nos níveis 1 a 4 foram praticamente idênticos nas duas análises.

**Palavras-chave:** *Transformada de ondaletas, Genoma, Zika vírus.*

---

<sup>1</sup>Departamento de Estatística, UFLA, E-mail: [leilamaria2003@yahoo.com.br](mailto:leilamaria2003@yahoo.com.br)

<sup>2</sup>Departamento de Estatística, UFLA, E-mail: [kelly\\_limaadm@hotmail.com](mailto:kelly_limaadm@hotmail.com)

<sup>3</sup>Departamento de Estatística, UFLA, E-mail: [safadi@des.ufla.br](mailto:safadi@des.ufla.br)

<sup>4</sup>Agradecimentos a Capes e Fapemig pelo apoio financeiro.

## Métodos de estimação para o expoente de Hurst

Aline Edlaine de Medeiros <sup>1 2</sup>

Mônica Marin de Souza <sup>3</sup>

Eniuce Menezes de Souza <sup>4</sup>

Uma das principais estatísticas utilizadas na análise de séries temporais com memória longa é um expoente de Hurst ( $H$ ). Esta estatística permite identificar a intensidade da autocorrelação presente na série temporal. A medida  $H \in (0, 1)$ , em particular, em particular, uma série temporal com  $H = 0.5$  representa o ruído branco. Diversas técnicas têm sido desenvolvidas para estimar o expoente de Hurst, tais como o Método da Variância Agregada, o Método de Higuchi, a Estatística R/S, o Método do Periodograma, entre outros. Neste trabalho, pretende-se avaliar a estimação do expoente de Hurst em dois tipos de séries temporais: um Ruído Gaussiano Fracionado e um Movimento Browniano Fracionado a partir de métodos empíricos. Tais séries temporais foram simuladas em linguagem R, onde também foram realizadas as estimativas de  $H$ . Foram simuladas séries temporais de tamanho 128, 512 e 2048, a partir dos valores 0.55, 0.75 e 0.95 para o expoente de Hurst. Observou-se que métodos distintos fornecem estimativas muitas vezes bem diferentes. Além disso, determinar o parâmetro de corte para o qual a aproximação linear possua um coeficiente angular que estime adequadamente  $H$ , não é uma tarefa fácil, sendo o método gráfico um dos recursos mais utilizados. Nota-se que estimativas obtidas graficamente para o parâmetro  $H$  podem ser bastante controversas, neste sentido, os métodos empíricos devem ser utilizados com cautela, e/ou combinado com outras técnicas, pois uma análise gráfica equivocada, ou ainda a não realização dela, podem levar a constatações espúrias sobre os dados.

**Palavras-chave:** *parâmetro de autossimilaridade, séries temporais, parâmetro de corte.*

---

<sup>1</sup>PBE - UEM. e-mail: [aline\\_medeiros@outlook.com](mailto:aline_medeiros@outlook.com)

<sup>2</sup>Agradecimento ao CAPES pelo apoio financeiro.

<sup>3</sup>PBE - UEM.

<sup>4</sup>DES - UEM



## Pacote BioBayes para o *software* R

Stéfane Lele <sup>1 3</sup>

Robson M. Rossi <sup>2 3</sup>

Este trabalho tem por objetivo a divulgação do pacote que está sendo construído no *software* R denominado BioBayes. A ideia principal é facilitar as análises bayesianas junto ao *software* OpenBUGS, em que os modelos no formato .bug já encontram-se implementados no pacote BioBayes. A coleção de funções de distribuições disponíveis estão listadas no *help* da função principal BioBayes.fit. Após as análises, um relatório é gerado automaticamente, contendo gráficos a respeito das cadeias, estimativas, densidades a posteriori para os parâmetros, análises de convergência, DIC, e o modelo utilizado no formato .bug. Este pacote ainda está em construção, portanto, não está disponível no repositório CRAN. A previsão para torná-lo disponível é neste ano (2018).

**Palavras-chave:** *Bayesiana, OpenBUGS, Posteriori.*

---

<sup>1</sup>DES - UEM. e-mail: [ste.tistics@gmail.com](mailto:ste.tistics@gmail.com)

<sup>2</sup>DES - UEM.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

# Ecologia

## Análise comportamental de ovelhas submetidas a intervenção humana usando GAMLSS

Lineu Alberto Cavazani de Freitas <sup>1</sup>

Cesar Augusto Taconeli <sup>2</sup>

José Luiz Padilha da Silva <sup>3</sup>

O presente trabalho teve como objetivo analisar o efeito da escovação no comportamento de ovelhas. Os dados foram coletados em um experimento com 20 animais classificados como reativos ou não ao isolamento social e avaliados em 3 sessões experimentais, e em 3 momentos distintos: antes, durante e após a escovação. A análise foi realizada por meio de Modelos Generalizados Aditivos para Localização, Escala e Forma (GAMLSS). Incluíram-se dois efeitos aleatórios: efeito aleatório de animal e de animal dentro de sessão. A inserção desses efeitos deve-se à necessidade de incorporar as correlações entre as medidas num mesmo animal e do animal numa mesma sessão. A análise foi realizada no software R, pacote gamlss, e duas variáveis foram consideradas: número de mudanças de postura de cauda e número de mudanças de postura de orelha, ambas variáveis de contagem. As distribuições consideradas foram: Poisson, Binomial Negativa e as suas versões inflacionadas de zeros e zero ajustadas. Modelou-se também o parâmetro associado à inflação de zeros em função dos fatores experimentais. Os resultados mostraram frequência maior de mudanças de postura de cauda durante a escovação; no entanto, para as mudanças de postura de orelha, menores frequências foram verificadas durante a escovação. O diagnóstico dos modelos, baseado em resíduos quantílicos aleatorizados, mostrou que os mesmos se ajustaram satisfatoriamente.

**Palavras-chave:** *Comportamento Animal, Dados de Contagem, GAMLSS.*

---

<sup>1</sup>PET-Est. UFPR. e-mail: [lialcafre@gmail.com](mailto:lialcafre@gmail.com)

<sup>2</sup>DEST - UFPR.

<sup>3</sup>DEST - UFPR.

## Definição de índices de competitividade de espécies nativas da Floresta Nacional de Carajás utilizando análise de fatores

Kaléo Dias Pereira<sup>1 3</sup>

Rafael Gomes Viana<sup>2 3</sup>

Cíntia Helena Marega<sup>2 3</sup>

Túlio Wanderson Ferreira Nunes<sup>2 3</sup>

Roberthi Alef Costa Teixeira<sup>2 3</sup>

O objetivo do presente trabalho foi utilizar a análise de fatores para definir índices de capacidade competitiva de três espécies nativas da Floresta Nacional de Carajás em competição com diferentes densidades de plantas da gramínea *Urochloa brizantha*. O experimento foi realizado em casa de vegetação e consistiu no plantio em vasos das espécies nativas *Bauhinia longipedicellata*, *Mimosa acutistipula* e *Solanum crinitum* em competição com a gramínea exótica *Urochloa brizantha*. A gramínea exótica foi estabelecida em densidades variando de 1 a 5 plantas por vaso com quatro repetições delineadas totalmente ao acaso. Foram avaliados parâmetros morfológicos e fisiológicos das plantas nativas, além de parâmetros morfológicos da gramínea exótica, totalizando em 16 variáveis mensuradas. Os dados foram submetidos à análise de fatores pelo método da máxima verossimilhança com rotação varimax. Na seleção dos fatores foram extraídos aqueles com autovalores maiores ou iguais a 1,0. As variáveis originais foram consideradas altamente correlacionadas aos fatores quando as cargas fatoriais assumiram valores  $\geq 0,6$ . A adequabilidade da análise de fatores foi avaliada aplicando-se o teste de esfericidade de Bartlett ao nível de 0,05 de significância. Constatou-se que apenas os primeiros quatro fatores foram associados a autovalores maiores que 1,0. Os quatro primeiros fatores explicaram 74% da variância dos dados e, com base no teste de esfericidade Bartlett, a análise de fatores mostrou-se adequada. A partir das cargas fatoriais, foi possível interpretar três dos quatro fatores extraídos, os quais foram definidos como índice de crescimento vegetativo, índice densidade de infestação e índice de qualidade fisiológica.

**Palavras-chave:** *Análise multivariada, Competição interespecífica, Recuperação de áreas degradadas.*

<sup>1</sup> DET – UFV. e-mail kaleoef@gmail.com

<sup>2</sup> ICA – UFRA.

<sup>3</sup> Agradecimentos à Vale S.A. pelo suporte logístico e financeiro.

## Retrospective estimation of growth of Chilean common hake

Cristian Villegas<sup>1</sup>

Emmanuel Lesaffre<sup>2</sup>

Víctor Espejo<sup>3</sup>

The age-at-capture and length-at-capture of common hake (*Merluccius gayi gayi*) was collected in San Antonio and Talcahuano areas in Chile between 1967 and 2010. In addition, the number of annual rings in sagittal otolith, a common estimate of the age of the fish, is determined. This estimate is, however, interval-censored. Research questions such as the comparison of the age-at-capture over the years and the relationship of age with length-at-capture require appropriate modeling techniques. It was also of interest to know whether there are different growth patterns inferred from the otolith radii in the hake and whether there is a trend in growth over the many years data were collected. This involves mixture modeling of longitudinal data. Because of the complexity of the data, a Bayesian approach was used to address the research questions.

**Keywords:** *Bayesian approach, Growth models, Interval-censored.*

---

<sup>1</sup>Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, Brazil. e-mail: *clobos@usp.br*

<sup>2</sup>L-Biostat, KU Leuven, Leuven, Belgium

<sup>3</sup>Subsecretaría de Pesca y Acuicultura, Valparaíso, Chile

## Uma abordagem para a análise de biomassa inflacionada de zeros

Júlio César Pereira<sup>1</sup>

Giovani Loiola da Silva<sup>2</sup>

Victória Isaac<sup>3</sup>

Em geral os dados de captura e esforço gerados por pescarias comerciais são os dados mais comuns disponíveis para avaliação de estoques. Este estudo foi motivado pela dificuldade em analisar dados das pescarias no rio Amazonas Baixas devido ao fenômeno de inflação zero. O objetivo deste trabalho foi desenvolver um modelo capaz de acomodar a inflação zero nas capturas em peso, permitindo uma melhor compreensão das variações da mesma relacionadas às variações no esforço e outras covariáveis disponíveis. Para analisar esses dados, foi desenvolvido um modelo hierárquico Bayesiano em três estágios, implementado no software OpenBugs. Na primeira etapa, modelou-se o número de viagens de pesca por local ( $N$ ), de acordo com uma distribuição de Poisson. No segundo estágio, dado  $N > 0$ , definiu-se uma variável Bernoulli  $X$ , onde  $X$  assume 1 se as capturas ocorreram para uma determinada espécie, e zero, se nada foi capturado para essa espécie. Finalmente, na terceira etapa, foi modelado o peso denotado por  $Y$ , que assume zero quando  $N = 0$  ou quando  $X = 0$  e  $N > 0$ . Quando  $X = 1$  e  $N > 0$  o peso foi assumido seguir distribuição gama, onde a média foi proporcional ao número de visitas  $N$ . Esta abordagem forneceu uma ferramenta útil para analisar a variação na captura por unidade de esforço como função de covariáveis quando os dados são inflados por zeros provenientes de ambas as fontes: abstinência da atividade pesqueira e ausência de captura em a presença de atividade pesqueira.

**Palavras-chave:** *Inflação de zeros, Modelo Poisson composto, Pesca.*

---

<sup>1</sup> DCA – UFSCAR. e-mail [julio.pereira.ufscar@gmail.com](mailto:julio.pereira.ufscar@gmail.com)

<sup>2</sup> IST – Universidade de Lisboa.

<sup>3</sup> Lab. de Biologia Pesqueira - UFPA.

# Um modelo flexível para dados de contagem não equidispersos: Uma aplicação em dados ecológicos com subdispersão

Douglas Toledo <sup>1 4</sup>

Cristiane Akemi Umetsu <sup>2</sup>

Idemauro Antonio Rodrigues de Lara <sup>3 4</sup>

Em pesquisas ecológicas é frequente o uso de dados de contagem como variável resposta. O modelo de regressão Poisson aparece como candidato principal para esse tipo de variável, entretanto, o modelo Poisson exige a equidispersão, isto é, média e variância iguais. No entanto, essa relação nem sempre ocorre, podendo a variância ser maior ou menor do que a média, fenômenos conhecidos como superdispersão e subdispersão, respectivamente. A não equidispersão, quando desconsiderada, pode acarretar em desvios padrões imprecisos, promovendo um modelo de ajuste inadequado. Neste contexto, esse trabalho apresenta o modelo COM-Poisson (CP) como um modelo mais flexível para problemas de contagem associados a não equidispersão, uma vez que o modelo CP apresenta um parâmetro adicional que permite maior flexibilidade. Como motivação, tem-se um estudo que provém de um trabalho de campo realizado na bacia hidrográfica do Rio Itanhaém, que objetivou avaliar a influência do aumento da biomassa de uma espécie exótica de capim aquático *Urochloa arrecta* sobre o número de grupos funcionais da comunidade de macrófitas nativas. A estimação dos parâmetros foi realizada via máxima verossimilhança e o parâmetro de dispersão foi testado por meio do teste de razão de verossimilhança, indicando a presença de subdispersão. Para a validação do modelo utilizou-se o gráfico *half-normal plot*. O modelo CP é paramétrico, sendo assim, foi possível calcular medidas de probabilidade para a distribuição das contagens. O modelo sugere que o crescimento da espécie exótica, expresso em biomassa, inibe o desenvolvimento de outros grupos funcionais.

**Palavras-chave:** *COM-Poisson*, *Urochloa arrecta*, *Poisson*, *half-normal plot*, *superdispersão*.

---

<sup>1</sup>LCE-ESALQ/USP. e-mail: [douglasbene@hotmail.com](mailto:douglasbene@hotmail.com)

<sup>2</sup>CAUNESP-UNESP.

<sup>3</sup>LCE-ESALQ/USP.

<sup>4</sup>Agradecimento ao CNPq pelo apoio financeiro.

# Economia



---

**ABERTURA FINANCEIRA, FLUXOS DE CAPITAIS E O COMPORTAMENTO DA TAXA DE CÂMBIO REAL NA ECONOMIA BRASILEIRA CONTEMPORÂNEA****Michael Gonçalves da Silva<sup>1 4</sup>****Vanessa Siqueira Peres da Silva<sup>2 4</sup>****Fernanda Dalcin Flôres<sup>3 4</sup>*****Resumo***

Para tentar analisar a volatilidade dos fluxos de capitais foi feita a estimação do modelo ARCH para cada série. Inicialmente estimou-se um modelo ARMA adequado à série original dos dados, visando remover, caso exista, a correlação serial da série. Foi verificado se a série original apresentou heterocedasticidade condicional. A partir da análise da função de autocorrelação (fac) e função de autocorrelação parcial (facp) escolheu-se o modelo ARIMA através de sua estrutura de parâmetros autorregressivos e de médias móveis. Em linhas gerais, os valores da fac significativos correspondem aos parâmetros de médias móveis, e os valores da facp significativos correspondem aos parâmetros da parte autorregressiva do modelo ARMA. Em seguida, aplicou-se a modelagem ARCH aos resíduos do modelo ARMA. Entretanto, foi necessário analisar a função de autocorrelação parcial (facp) dos quadrados do resíduo do modelo ARMA. Os *lags* das autocorrelações significativas dessa função indicaram os parâmetros autorregressivos do modelo ARCH. Posteriormente, as propriedades dos resíduos do modelo ARCH identificados foram analisados, verificando a condição de normalidade, homocedasticidade e não correlação. E, finalmente, previu-se a variância condicional, ou seja, a volatilidade do modelo da série em questão. Os Vetores Autorregressivos foram estimados para serem geradas as Funções de Impulso e Resposta e a decomposição de variância. Assim, dentre os fluxos de capitais selecionados, constatou-se que os investimentos em carteira, tanto estrangeiros quanto brasileiro, tendem a apreciar a taxa de câmbio real da economia brasileira entre os anos de 2000-2013.

---

<sup>1</sup> UNIFRA, e-mail michael.silva@ufsm.br

<sup>2</sup> CCNE - UFSM.

<sup>3</sup> CCSH – UFSM

<sup>4</sup> Agradecimento ao CNPq pelo apoio financeiro.

## **A mortalidade infantil no estado de São Paulo: uma previsão por meio da modelagem SARIMA**

**Diana Chaukat Chaib<sup>1</sup>**

**Taylor Oliveira Fidelis<sup>2</sup>**

**Walef Machado de Mendonça<sup>3</sup>**

**Manoel Vitor Souza Veloso<sup>4</sup>**

O índice de mortalidade infantil tem se tornado, sobretudo nas últimas décadas, um dos principais indicadores do nível de vida da população. De acordo com Singer (1974), a mortalidade infantil se enquadra mais como um indicador de padrão de vida do que saúde, uma vez que, a sobrevivência de crianças menores que um ano depende mais de condições de nutrição e higiene do que de cuidados médicos. Desta forma, a pesquisa tem por objetivo, através das técnicas estatísticas e econométricas, estabelecer previsões, com intervalo de confiança de 5%, da mortalidade infantil para o estado de São Paulo. Foi utilizado o modelo autorregressivo integrado de médias móveis (ARIMA), uma vez que este explica uma variável através de valores passados. Esse modelo foi proposto por Box e Jenkins na década de 1970 e tem origem nos modelos autorregressivo (AR), médias móveis (MA) e da combinação dos modelos AR e MA (ARMA). Além de incluir modelos sazonais (SARIMA). Os dados para a análise foram mensais do período de Janeiro de 1996 a Dezembro de 2016. Foram coletados na página eletrônica do Sistema IBGE de Recuperação Automática (SIDRA). Os dados foram rodados no software livre R Studio e a previsão foi feita para os próximos cinco meses subsequentes ao período analisado. A previsão do comportamento da variável aproximou-se das teorias utilizadas como base, uma vez que ambas indicam que a mortalidade infantil vem nos últimos anos apresentando declínio. A pesquisa visou contribuir para as análises acerca da mortalidade infantil por meio de modelos econométricos.

**Palavras-chave:** *Mortalidade infantil, previsão, SARIMA.*

---

<sup>1</sup> UNIFAL - dianachaib@hotmail.com

<sup>2</sup> UNIFAL- taylorfidelis@hotmail.com

<sup>3</sup> UNIFAL – [walefm2@gmail.com](mailto:walefm2@gmail.com)

<sup>4</sup> UNIFAL – manoel.souza@unifal-mg.edu.br

# Análise bayesiana do prêmio de seguro em função da quantidade de roubos de automóveis no Brasil

Josiane Correia de Souza Carvalho <sup>1 4 5</sup>

Luiz Otávio de Oliveira Pala <sup>2 4 5</sup>

Manoel Vitor de Souza Veloso <sup>3 5</sup>

Os prêmios são valores pagos as seguradoras, para que sejam responsáveis pelos efeitos econômicos de algum sinistro inerente ao bem segurado. O valor definido nas apólices de seguro para tais prêmios variam anualmente, e tal variação é consequência de alguns fatores que estão diretamente ligados aos contratos de seguro. Dentre os seguros no ramo não-vida, o de automóveis é um dos mais comercializados no Brasil. Desta forma, o objetivo deste trabalho é analisar a variação do prêmio em função do roubo de veículos. Para tal objetivo, foi feita uma análise bayesiana para o ajuste de um modelo de regressão, utilizando o software MultiBUGS, considerando dados dos 150 principais modelos de veículos para os últimos três períodos (anos de 2008, 2009 e 2010) disponíveis no Brasil. O diagnóstico de Gelman e Rubin indicou que houve convergência das cadeias para os parâmetros estimados. Foram calculadas as estimativas *a posteriori* por ponto (média) e por intervalo (IC-HPD com 95%) para os parâmetros do modelo, as quais indicaram que, com o aumento em 1% no número de roubos o prêmio aumenta 0,56%, o que, atuarialmente, permite levantar a questão de solvência neste setor, sob a ótica do impacto nos prêmios dado o número de roubos. Para o período em questão, evidenciou que há, com o passar dos anos, um aumento nos prêmios, de forma que no ano de 2008 eram 12,37% menores quando comparados ao ano de 2010, e eram 8,5% menores em 2009, quando comparado com esse mesmo ano. Tais resultados conduzem a uma discussão sobre os métodos para o gerenciamento de risco que, quanto melhor ajustados, auxiliam no equilíbrio atuarial.

**Palavras-chave:** *Inferência Bayesiana. Seguros. Prêmio.*

---

<sup>1</sup>Programa de Pós Graduação em Estatística Aplicada e Biometria - UNIFAL-MG e-mail: [josianecorreia.carvalho@gmail.com](mailto:josianecorreia.carvalho@gmail.com)

<sup>2</sup>Programa de Pós Graduação em Estatística Aplicada e Biometria - UNIFAL-MG.

<sup>4</sup>Agradecimento à CAPES pela bolsa.

<sup>3</sup>Instituto de Ciências Sociais Aplicadas- UNIFAL-MG

<sup>5</sup>Agradecimento à FAPEMIG pelo apoio financeiro

# Análise da Exportação de Café Brasileiro no período de 1994 a 2010

Fernanda Mayumi Nishihata <sup>1 4</sup>

Leonardo Henrique de Andrade B. do Nascimento <sup>2 4</sup>

Ana Carolina de Oliveira <sup>3 4</sup>

Nas últimas décadas é de grande notoriedade a forte redução da participação do café brasileiro na receita de divisas e no comércio internacional. No entanto, o café ainda desempenha um papel relevante na economia nacional. Desse modo, o objetivo deste trabalho é avaliar qual o impacto de variáveis macroeconômicas sobre as variações ocorridas na exportação de café brasileiro (sacas 60kg). Para isso utilizou-se uma série temporal anual de 1994 a 2010, retirados do banco de dados *IpeaData*. Utilizou-se o indexador inflacionário IGP-DI para fins de atualização monetária. A abordagem escolhida para a análise dos dados foi de natureza quantitativa com a utilização de Regressão Linear Múltipla (RLM) ajustada pelo método dos Mínimos Quadrados Ordinários (MQO) através do *software* gratuito *Gretl* 1.9.90. Os pressupostos acerca da RLM foram testados e respeitados. Em relação ao modelo, a primeira análise mostrou que a variável *Produção* não era significativa. Desta forma, o modelo reduzido contou com as variáveis independentes IPCA, PIB e Taxa de Câmbio. O coeficiente de determinação foi de 0,9139, ou seja, das variações ocorridas na Exportação, 91,39% delas são explicadas pelas oscilações nas variáveis independentes. Essa estatística foi comprovada mediante a realização do teste *F* que se mostrou significativo ao nível de 1%. O coeficiente de todas as variáveis pertencentes ao modelo reduzido foram consistentes, o que é comprovado pelo teste *t*, através do qual todas se mostraram significativas ao nível de 1%. Além disso as variáveis IPCA, PIB e Taxa de Câmbio, apresentam um relacionamento positivo com a variável dependente, corroborando com as hipóteses iniciais.

**Palavras-chave:** *Exportação de Café, Regressão Linear Múltipla, Mínimos Quadrados Ordinários.*

---

<sup>1</sup>ICSA - UNIFAL-MG. e-mail: [fernandamayumi@yahoo.com.br](mailto:fernandamayumi@yahoo.com.br)

<sup>2</sup>ICSA - UNIFAL-MG. e-mail: [leonardoandrade.h@outlook.com](mailto:leonardoandrade.h@outlook.com)

<sup>3</sup>ICSA - UNIFAL-MG. e-mail: [ana.coliveira@outlook.com](mailto:ana.coliveira@outlook.com)

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

# Análise das cotações da commodity do café brasileiro utilizando o modelo Bayesiano de volatilidade estocástica SV-AR(p)

Luiz Otávio de Oliveira Pala<sup>1 3</sup>

Gislene Araujo Pereira<sup>2 3</sup>

Letícia Lima Milani Rodrigues<sup>2 3</sup>

O mercado de commodities tem possibilitado a expansão econômica de países, principalmente quanto as formas de negociação e níveis de venda. No contexto econômico brasileiro, nota-se a importância destas mercadorias para o país, associando-se ao nível de desenvolvimento das exportações e servindo como opção para investidores. As commodities cafeeiras são consideradas as mais comercializadas internacionalmente, sendo alvos de estudos quanto ao comportamento de seus preços. Este trabalho teve por objetivo a análise da cotação internacional do café brasileiro em centavos de dólar americano, durante o período de 1964 a 2016, sob a forma de série temporal mensal disponibilizada pelo Instituto de Pesquisa Econômica Aplicada. Para tanto, utilizou-se da análise de modelos de séries temporais de volatilidade estocástica, via modelo SV-AR(p) sob o enfoque Bayesiano, buscando a exploração de níveis de oscilação das cotações em períodos de incertezas econômicas no mercado e realizando previsões para a volatilidade futura. Tal análise foi realizada no *Software R*, possibilitando a estimação do modelo e análise. Os resultados indicaram que a estrutura de variância das cotações de café apresentaram períodos de maior volatilidade durante a década de 1970 e 1980, resultando também em previsões da volatilidade para o ano de 2017. Desta forma, este trabalho contribui quanto a exploração de modelos de volatilidade estocástica em cotações de commodities, facilitando as estratégias de mercado.

**Palavras-chave:** *Café Brasil, Commodities, Volatilidade.*

---

<sup>1</sup>Graduando em Ciências Atuariais. ICSA - Unifal. e-mail: [luizotavio.oliveira@gmail.com](mailto:luizotavio.oliveira@gmail.com).

<sup>2</sup>Professora adjunta ICSA - Unifal.

<sup>3</sup>Agradecimento ao CNPq, CAPES, FAPEMIG e UNIFAL pelo apoio financeiro.

## **Análise da Taxa Bruta de Mortalidade para o estado de Minas Geras utilizando Regressão Linear Múltipla**

Karoline Sanchez Fernandes<sup>1 3</sup>

Mayara Cristine dos Reis<sup>1 3</sup>

Jayne de Proença Salles<sup>1 3</sup>

Manoel Vitor de Souza Veloso<sup>2 3</sup>

O presente trabalho tem como objetivo explicar a Taxa Bruta de Mortalidade (TBM) do estado de Minas Gerais, no ano de 2010, em função de 9 variáveis macroeconômicas, dentre elas o Índice de GINI, que mede o grau de desigualdade existente na distribuição de renda dos indivíduos e outras variáveis associadas ao saneamento básico e renda populacional. A análise foi realizada por meio de uma Regressão Linear Múltipla (RLM). Os dados foram coletados dos relatórios fornecidos pelo Programa das Nações Unidas para o Desenvolvimento – PNUD. Utilizou-se do procedimento de *stepwise* para seleção de variáveis, obtendo um modelo com três variáveis independentes sendo elas RDPC (razão entre o somatório da renda de todos os indivíduos residentes em domicílios particulares permanentes e o número total desses indivíduos;), GINI e PIND (proporção dos indivíduos com renda domiciliar per capita igual ou inferior a R\$ 70,00 mensais) e seus coeficientes significativos ao nível de 5%. Na análise de resíduos foi constatado que seguem uma distribuição normal, são independentes e homocedásticos. Desta forma, percebe-se que a economia não se preocupa unicamente com a renda e com a riqueza, mas sim como esses recursos podem ser melhor aproveitados. Os resultados obtidos mostram a viabilidade de relacionar estatísticas de mortalidade com o grau da desigualdade de distribuição de renda nos municípios do estado de Minas Gerais. Logo, este trabalho contribui como uma forma de visualizar que a TBM possui uma relação com a carência econômica, e que esta afeta diretamente a população envolvida.

**Palavras Chaves:** Mortalidade, Regressão Linear Múltipla, Mínimos Quadrados Ordinários.

---

<sup>1</sup> ICSA– UNIFAL-MG. e-mail: karol-sanchez@live.com; mayara\_creis@yahoo.com.br; jayne-ja@hotmail.com.

<sup>2</sup> Professor adjunto ICSA – UNIFAL-MG.

<sup>3</sup> Agradecimento ao CNPq, CAPES, FAPEMIG e UNIFAL pelo apoio financeiro.

## Análise da volatilidade do retorno de Bitcoins

Jaqueline Gonçalves Fernandes<sup>1 3</sup>

Marcela de Marillac Carvalho<sup>2 3</sup>

Kelly Pereira de Lima<sup>2 3</sup>

Thelma Sáfadi<sup>2 3</sup>

A palavra Bitcoin pode ser dividida em duas outras palavras, *bit* que corresponde a dígito binário e *coin* que significa moeda em inglês. Assim, o Bitcoin pode ser considerado como uma espécie de dinheiro, porém na forma digital e totalmente descentralizado. A forma como é determinada o valor da moeda depende exclusivamente do comportamento dos indivíduos que estão envolvidos nesse mercado. Ulrich (2017) compara o Bitcoin com a internet, ele relaciona a agilidade e eficácia de como são repassadas as mensagens com o que essa nova moeda irá fazer com o dinheiro. Diante desse cenário e do interesse de várias pessoas em entrar nesse mercado, torna-se necessário estudar e entender o comportamento da valorização dessa moeda ao longo do tempo. Recorre-se aos processos de séries temporais referentes às classes ARFIMA(p,d,q)-GARCH(1,1), ARFIMA(p,d,q)-APARCH(1,1), ARFIMA(p,d,q)-EGARCH(1,1) e ARFIMA(p,d,q)-IGARCH(1,1) com a distribuição de erros t-skewed. Pelo critério de AIC e com a realização de diagnósticos nos resíduos, por meio do teste Box-Pierce, o modelo adequado a série foi um ARFIMA(1,d,1)-EGARCH(1,1), sendo d uma diferenciação de ordem fracionária com  $d=0,22$ . Os resultados indicam um comportamento de memória longa nos retornos dos bitcoins, bem como a existência de efeito da alavancagem ao longo do período, ou seja, choques negativos passados têm maior impacto sobre a volatilidade condicional no presente. Ademais, pode-se observar uma persistência de choques passados na volatilidade da série de retorno do bitcoins por vários períodos.

**Palavras-chave:** AFIRMA, Criptomoeda, Séries temporais.

---

<sup>1</sup> DES – UFLA. e-mail: jaqgonfer@gmail.com

<sup>2</sup> DES – UFLA.

<sup>3</sup> Agradecimento a CAPES, CNPq e FAPEMIG pelo apoio financeiro.

---

## **Análise do consumo de gasolina no Brasil do período de 2005 a 2017 a partir da metodologia de Box-Jenkins**

**Bárbara Alves Lopes<sup>1 3</sup>**

**Manoel Vítor de Souza Veloso<sup>2 3</sup>**

**Letícia Lima Milani<sup>2 3</sup>**

A gasolina que é derivada do petróleo é considerada um dos compostos mais utilizados na indústria automobilística no Brasil e no mundo. Ao longo dos anos o consumo deste tipo de combustível vem sendo marcado por diversas fases tais como os choques relacionados ao petróleo, a criação do Programa Nacional do Alcool (PNA), o avanço tecnológico do modelo de automóveis flex que permite a escolha entre álcool e gasolina afetando assim a demanda da mesma a partir de 2005. O presente trabalho tem como objetivo a análise da série temporal do consumo aparente gasolina em barril por mês no período de 2005 a 2017, utilizando a metodologia de Box e Jenkins (1970). Para a realização deste, foram realizados testes de Cox-Stuart para a tendência e de Fisher para a sazonalidade, confirmando assim a presença de ambas. No intuito de tornar a série estacionária, aplicou-se o método da diferenciação. O teste de raiz unitária foi aplicado, confirmando que a série diferenciada é estacionária. O melhor modelo ajustado foi escolhido com base no critério de Akaike, sendo este o  $SARIMA(2,1,2)(0,1,1)_{12}$ , apropriado também para realizar as previsões.

**Palavras-chave:** *Box-Jenkins, consumo de gasolina, séries temporais.*

---

<sup>1</sup> ICSA – UNIFAL. e-mail barbaraalvesls@yahoo.com.br

<sup>2</sup> ICSA - UNIFAL

<sup>3</sup> Agradecimento à FAPEMIG pelo apoio financeiro.



# Análise do desempenho bancário brasileiro utilizando análise multivariada e metodologia CAMEL

Taylor Oliveira Fidelis <sup>1 3</sup>

Gabriel Rodrigo Gomes Pessanha <sup>2 3</sup>

<sup>3</sup>

O presente trabalho se propõe a analisar e comparar o desempenho das instituições bancárias brasileiras entre si no recorte temporal entre janeiro de 2008 até junho de 2017. Para avaliação das instituições adotou-se as variáveis segundo a metodologia CAMEL que contempla uma visão abrangente de análise financeira por meio de cinco dimensões diferenciadas: C- Capital, A-Qualidade dos Ativos, M-Gerenciamento, E-Lucratividade e L- Liquidez. Já a avaliação da performance de cada instituição frente ao setor Bancário deu-se via componentes principais e aplicação de métodos de Agrupamento, e por fim ranking dos bancos ao longo do tempo. Todas as análises e tratamento dos dados foram realizados por meio do software livre RStudio, quanto aos balancetes ou dados contábeis mensais foram obtidos pela plataforma do Banco Central do Brasil. Na fundamentação teórica foram estudadas as técnicas multivariadas, técnicas de avaliação de instituições financeiras em específico bancos, teses e artigos que empregam outras metodologias de avaliação. Os resultados obtidos evidenciaram a eficácia do método via análise multivariada para avaliação no desempenho financeiro junto a metodologia CAMEL. Quanto a pertinência deste estudo, traduz-se na necessidade tanto dos órgãos regulamentadores como das próprias instituições financeiras, buscarem de forma eficiente e rápida avaliar o desempenho temporal de uma dada instituição mediante ao comportamento do mercado setorial.

**Palavras-chave:** *Indicadores Financeiros, Análise multivariada, Componentes principais, Análise de agrupamento, Ranking, CAMEL .*

---

<sup>1</sup>Universidade Federal de Alfenas - UNIFAL-MG. email:taylorfidelis@hotmail.com

<sup>2</sup>Universidade Federal de Alfenas - UNIFAL-MG. email:gabrielrgp@yahoo.com.br

<sup>3</sup>Agradecimento à CNPQ e FAPEMIG pelo apoio financeiro.

## Análise do impacto do consumo das famílias na taxa de crescimento econômico utilizando o modelo VAR

Isabella Oliveira Martins <sup>1 3</sup>

Manoel Vitor de Souza Veloso <sup>2 3</sup>

A teoria econômica estabelece que o consumo é uma variável de grande peso na determinação da taxa de crescimento econômico. Dessa forma, o presente trabalho tem como objetivo analisar o impacto de um choque no consumo das famílias sobre a taxa de crescimento real da economia durante o terceiro trimestre de 1999 até o terceiro trimestre de 2015. Para tal, é feito o ajuste de um modelo VAR (1), o cálculo da função de resposta do crescimento econômico ao impulso no consumo das famílias e a decomposição da variância do PIB. O modelo VAR (vetor autorregressivo) considera todas as variáveis como endógenas e analisa as relações lineares existentes entre as variáveis e seus valores defasados no intuito de responder qual será a trajetória da série dado um choque estrutural. Para um melhor entendimento dessas inter-relações, utilizam-se como ferramentas a função de resposta ao impulso e a decomposição da variância. Foram analisados dados trimestrais referentes a variação da despesa com consumo das famílias, calculada pelo IBGE e a taxa de crescimento real do PIB, também disponibilizada pelo IBGE, totalizando 65 observações de cada variável. A análise da decomposição da variância mostra que o consumo é responsável por aproximadamente 32% da variância do PIB. Já o resultado obtido pela função de resposta ao impulso indica que, ao contrário do esperado, um choque no consumo causa uma queda significativa a 5% no PIB nos dois primeiros períodos. Dessa forma, conclui-se que políticas governamentais voltadas para o estímulo do consumo das famílias não são favoráveis.

**Palavras-chave:** *Consumo das famílias, Crescimento do PIB, Modelo VAR.*

---

<sup>1</sup>ICSA - UNIFAL. e-mail: *isabella\_martins26@hotmail.com*

<sup>2</sup>ICSA - UNIFAL.

<sup>3</sup>Agradecimento ao CNPq, FAPEMIG e CAPES pelo apoio financeiro.

## **Análise do Índice Bovespa (Ibovespa) de 2009 a 2018 através do Modelo ARIMA**

**Beatriz de Souza Bernardino<sup>1</sup>**

**Luiz Otávio de Oliveira Pala<sup>2</sup>**

**Gislene Pereira Araújo<sup>3</sup>**

**Kristopher Soares Caldeira<sup>4</sup>**

O Ibovespa é considerado o índice mais importante de performance, já que expressa o desempenho médio das cotações dos ativos de maior negociabilidade e representatividade dentro desse mercado. O mercado acionário brasileiro tem vantagens únicas em relação à rentabilidade de investimentos, entretanto, possui um risco considerável e precisa da oferta de ferramentas que auxiliem a minimizá-los (AMARO; SOUZA; 2014). Assim, o objetivo desse trabalho é avaliar o comportamento do Ibovespa no período de 03 de agosto de 2009 a 23 de janeiro de 2018 através da modelagem de Box-Jenkins, visando auxiliar os investidores a minimizar possíveis riscos futuros. Os dados foram coletados no site da *Yahoo Finance* e correspondem a informações diárias do preço de fechamento do índice, sendo tratados e ajustados no *software R*. Foram aplicados os testes de *Dickey Fuller* e *Cox Stuart* observando assim que a série era não estacionária e apresentava tendência. Assim, foi aplicada a primeira diferença à série e os testes foram refeitos, tendo como resultado uma série estacionária e sem tendência. Dessa maneira, com base nas funções ACF e PACF, o modelo resultante foi ARIMA(3,1,1), sendo o mais adequado devido ao nível de significância e o segundo menor AIC. Por fim, aplicou-se o teste de *Box-Pierce* e verificou-se que os resíduos são resíduo branco, com valor-p de 0,986. Comparando com os dados reais, o modelo obteve um bom comportamento diante da série.

**Palavras-chave:** *Box e Jenkins, Índice Bovespa, Risco.*

---

1 Graduanda em Ciências Atuariais – UNIFAL. E-mail: beatrizsb08@gmail.com

2 Mestrando em Estatística Aplicada e Biometria - UNIFAL

3 Professora Doutora do Instituto de Ciências Sociais Aplicado – ICSA/UNIFAL

4 Graduando em Bacharelado Interdisciplinar em Ciência e Economia - UNIFAL

## **Análise espacial do valor segurado por hectare das lavouras de café em MG**

**Jorge Garcia Filho<sup>1 4</sup>**

**Marcelo Silva de Oliveira<sup>2 4</sup>**

**Pablo Lourenço Ribeiro de Almeida<sup>3 4</sup>**

Atualmente o Brasil é o maior produtor e exportador de café do mundo, sendo Minas Gerais responsável por mais de 50% da produção nacional. Por meio das exportações do produto o país gera renda para o crescimento e fortalecimento da economia. Além de gerar receitas para o país, a atividade cafeeira é responsável pela geração de empregos direta ou indiretamente. Assim como qualquer outra atividade econômica, a atividade cafeeira carrega consigo alguns riscos relacionados à produtividade e comercialização. O seguro rural se mostra como um mecanismo de administração dos riscos relacionados à produção agropecuária, porém, ainda é um mecanismo pouco difundido e utilizado pelos produtores rurais, representando, em média, apenas 1% do mercado de seguros no Brasil. Dada a importância da atividade cafeeira para a economia do país, em especial para a economia mineira, e a necessidade do fortalecimento do mercado de seguro rural para uma maior proteção de nossas lavouras, o presente trabalho tem por objetivo analisar como os municípios mineiros asseguram espacialmente suas produções de café e como a proteção da lavoura se correlaciona espacialmente com a proteção dos municípios vizinhos. Para tanto será realizada uma análise espacial da autocorrelação do Valor Segurado por Hectare para os anos de 2015 e 2016 por meio da análise de dados de área utilizando o índice de autocorrelação I de Moran para captar e mensurar a espacialidade existente nos dados.

**Palavras-chave:** *Autocorrelação espacial, I de Moran, Seguro Rural*

---

<sup>1</sup> DES - UFLA. e-mail [jorgegarcia@posgrad.ufla.br](mailto:jorgegarcia@posgrad.ufla.br)

<sup>2</sup> DES - UFLA.

<sup>3</sup> DES - UFLA. e-mail [pablo\\_lourenco@hotmail.com](mailto:pablo_lourenco@hotmail.com)

<sup>4</sup> Agradecimento à CAPES, CNPQ e FAPEMIG pelo apoio financeiro.

## **Análise, por meio de componentes principais, dos indicadores econômicos que mais afetam a decisão dos investidores: um estudo de caso da Gerdau**

Isabella Oliveira Martins <sup>1 5</sup>

Marcela de Marillac Carvalho <sup>2 5</sup>

Luiz Otávio de Oliveira Pala <sup>3 5</sup>

Manoel Vitor de Souza Veloso <sup>4 5</sup>

A análise do desempenho das empresas, a partir dos indicadores financeiros, tem sido foco de estudo nos últimos anos. Nesse contexto, a Análise de Componentes Principais (ACP) foi utilizada para verificar quais os indicadores econômicos-financeiros da empresa Gerdau de maior relevância no período que compreende do primeiro trimestre de 2010 ao primeiro trimestre de 2017. Essa técnica multivariada consiste no agrupamento dos indivíduos de uma população segundo a variação de suas características, possibilitando representar os indicadores contábeis por componentes, independentes entre si, resultantes de adequadas combinações lineares das variáveis financeiras selecionadas, extraíndo-se o máximo de informação, em termos da variação total, contida nos dados e obtendo o grau de importância pelas correlações resultantes entre as componentes principais e as variáveis originais. Seis indicadores foram utilizados na análise sendo eles a margem EBTIDA que é o lucro antes dos juros, impostos, depreciação e amortização; o Lucro Por Ação (LPA), o Retorno Sobre o Ativo (ROA), o Retorno Sobre Capital Social (ROE), o Valor Patrimonial por Ação (VPA) e Vendas por Ação (VA). Os resultados obtidos pela ACP, por meio da matriz de correlações, apontam que os indicadores da empresa Gerdau podem ser resumidos a dois componentes principais que explicam aproximadamente 94% da variância total dos dados. Verifica-se, nesse trabalho, que os indicadores que se mostraram com maior relevância são os que mais afetam a tomada de decisão por parte dos investidores.

**Palavras-chave:** *Componentes Principais, Indicadores financeiros, Investimento.*

---

<sup>1</sup>ICSA - UNIFAL. e-mail: *isabella\_martins26@hotmail.com*

<sup>2</sup>DES - UFLA.

<sup>3</sup>ICEX - UNIFAL.

<sup>4</sup>ICSA - UNIFAL.

<sup>5</sup>Agradecimento ao CNPq, FAPEMIG e CAPES pelo apoio financeiro.

## **Análise temporal do número de endividamento e inadimplência das famílias no Brasil, no período de janeiro de 2010 a dezembro de 2017, utilizando a metodologia GLARMA Poisson**

**Camila Moreira Ribeiro**<sup>1 3</sup>

**Letícia Lima Milani**<sup>2 3</sup>

**Resumo:** O nível de endividamento e inadimplência do consumidor vem crescendo ao longo dos anos no mundo inteiro. Tal fato pode estar associado à baixa renda familiar, falta de planejamento financeiro e desemprego. Diante disto, o objetivo deste trabalho é fazer uma análise da série temporal do número de endividamento e inadimplência das famílias no Brasil no período de janeiro de 2010 a dezembro de 2017, buscando examinar a influência das covariáveis Taxa Selic, Índice de Preços ao Consumidor Amplo (IPCA) e Índice de Consumo das Famílias (ICF) aos dados. Para esse estudo foram realizadas análises usando a metodologia GLARMA Poisson, a fim de identificar o modelo mais adequado para a série em estudo. Por meio dos resultados, verificou-se que as covariáveis apresentadas estão relacionadas ao nível de endividamento e inadimplência do consumidor. Além disto, observou-se que o modelo GLARMA Poisson atendeu bem a modelagem, apresentando um bom ajuste a série de dados.

**Palavras-chave:** *Endividamento, GLARMA Poisson, Inadimplência.*

---

<sup>1</sup> ICSA – UNIFAL, email: camilam.ribeiro@yahoo.com.br

<sup>2</sup> Professora Adjunta UNIFAL-MG *campus* Varginha.

<sup>3</sup> Agradecimentos a CNPq e FAPEMIG pelo apoio financeiro.

---

## **Análise Temporal dos índices de desemprego no período de março de 2002 a fevereiro de 2016**

**Fábio Henrique de Jesus dos Santos<sup>1 3</sup>**

**Rafaela da Silva Gomes<sup>2 3</sup>**

**Luciene Resende Gonçalves<sup>2 3</sup>**

Desde o surgimento do capitalismo, o desemprego vem sendo motivo de preocupação de grande parte da população mundial, mesmo em países desenvolvidos como Estados Unidos e países da União Europeia, atingidos pelo desemprego principalmente após a crise mundial de 2008. Entretanto, o desemprego afeta de forma mais acentuada países subdesenvolvidos e emergentes, como o Brasil. Neste trabalho será utilizada a técnica de séries temporais na obtenção de valores futuros. As estimativas prévias dos índices de taxa de desemprego poderão ser utilizadas no planejamento do governo e podem ser de grande utilidade na atual conjuntura vivenciada no país no contexto político e econômico. O SARIMA (0,1,0) (2,1,0) com intervenções nas observações 40 e 83, referente respectivamente, a junho de 2005 e janeiro de 2009, foi o que se ajustou mais adequadamente aos dados e pode ser usado para fins de obtenção de valores futuros.

**Palavras-chave:** *Desemprego, SARIMA, Séries Temporais..*

---

<sup>1</sup> ICSA– Unifal. e-mail fabio\_h\_santos@hotmail.com

<sup>2</sup> ICSA– Unifal.

<sup>3</sup> Agradecimento a FAPEMIG pelo apoio financeiro.

## **Aplicação do MRLM para modelagem do gasto da população brasileira com gasolina comum**

**Zeferino Gomes da Silva Neto<sup>13</sup>**

**Danilo Régis dos Santos<sup>2</sup>**

**Carlos Ayallas de Sousa Pereira<sup>13</sup>**

**Valmária Rocha da Silva Ferraz<sup>2</sup>**

Com o crescente avanço da frota de automóveis, tem-se o aumento da necessidade de uso de combustíveis. A descoberta do petróleo implica no surgimento dos derivados do mesmo, como a gasolina e o óleo diesel. A gasolina no decorrer dos anos tem tido um papel fundamental nos transportes, não só no Brasil, mas também no mundo. Tendo em vista tal cenário a proposta de realização do trabalho aqui apresentado, tem como objetivo o estudo da relação do gasto dos brasileiros com gasolina comum, acarretado ao Produto Interno Bruto (PIB) brasileiro, e ao preço da gasolina comum utilizando uma metodologia estatística denominada análise de Regressão Linear Múltipla (RLM) no software estatístico R. O estudo foi realizado no período de Janeiro de 2011 a Janeiro de 2017. Os dados foram coletados trimestralmente, posteriormente, foram submetidos aos pressupostos do modelo de RLM, dentre as quais, Ausência de Outliers, Normalidade, Homocedasticidade, Independência e Multicolineariedade, foram verificadas e diagnosticadas, através disso, conseguimos generalizar o modelo para além da amostra, tal ferramenta de modelagem permite tirar estatísticas inferenciais. O modelo ajustado neste estudo explica aproximadamente 74,05% da variável de interesse, o consumo de gasolina comum, o modelo também permitiu a obtenção de valores futuros acerca do gasto da população brasileira com gasolina comum, os resultados mostram-se satisfatórios, ao comparar a predição do modelo com os valores reais do trimestre subsequente a amostra, pode-se constatar um erro de apenas 6,16% da estimativa para o valor real. Para estudos futuros, pretende-se implementar o modelo com novas variáveis explicativas.

**Palavras-chave:** *Gasolina, Previsão, Regressão Linear Múltipla.*

---

<sup>1</sup> DET – UFV. e-mail [zeferinon@gmail.com](mailto:zeferinon@gmail.com)

<sup>2</sup> UFPI

<sup>3</sup> Agradecimento à FAPEMIG e a CAPES pelo apoio financeiro.



## A previdência social nos municípios de Minas Gerais: uma tipologia utilizando análise multivariada

Walef Machado de Mendonça <sup>1 3</sup>

Ana Carolina Orrico <sup>2 3</sup>

Patrícia de Siqueira Ramos <sup>2 3</sup>

A previdência social no Brasil tem grande importância como uma das principais políticas sociais, sendo ela responsável pela manutenção da atividade econômica em municípios de pequeno porte (FERREIRA; SOUZA. 2008). O objetivo deste trabalho é a avaliação das principais variáveis socioeconômicas relacionadas à previdência por meio de dois métodos de análise multivariada: componentes principais (ACP) e agrupamento (AA), utilizando-se a linguagem *Python*. Os dados são provenientes do Ministério da Previdência e do IBGE. De acordo com a ACP, os três primeiros componentes explicam 71,86% da variância total. As variáveis com maiores coeficientes no CP1 foram: quantidade de benefícios per capita, valor dos benefícios per capita e PIB per capita. Para efetuar o agrupamento, utilizou-se o método de Ward para a definição do número de grupos. O ponto de corte do dendrograma foi escolhido a uma altura que dividiu os municípios brasileiros em quatro grupos. O método das *k*-médias resultou em grupos com número desbalanceado de municípios: 3.983, 1.445, 119 e 17. O grupo 1 é o que possui menor valor de arrecadação média per capita (R\$ 158,27), menor esperança de vida (72,68 anos) e menor valor médio do benefício per capita (R\$ 1.757,29) e o grupo 2 apresentou o maior valor de arrecadação per capita (R\$ 795,88). O grupo 3 possui a maior esperança de vida (75,17 anos). Além disso, um mapa foi gerado com a distribuição dos municípios nos grupos: os do grupo 1 concentram-se no Norte e Nordeste e os do grupo 2 no Centro-Oeste, Sudeste e Sul.

**Palavras-chave:** *Agrupamento, Componentes Principais, Municípios, Previdência Social.*

---

<sup>1</sup>ICSA/UNIFAL-MG e-mail: [walefm2@gmail.com](mailto:walefm2@gmail.com)

<sup>2</sup>ICSA/UNIFAL-MG

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

# AS COTAÇÕES DIÁRIAS DA JBS S/A EM 2017: MODELAGEM UTILIZANDO UMA ANÁLISE DE INTERVENÇÃO.

Raquel Maduro Ayres Bandeira <sup>1 3</sup>

Manoel Vitor de Souza Veloso <sup>2 3</sup>

Em maio de 2017 os executivos da JBS, foram alvo de investigações pelo Supremo Tribunal Federal e tiveram suas delações premiadas divulgadas pela imprensa. No intuito de estipular o desempenho eficiente do mercado financeiro, este estudo analisa as cotações diárias da empresa ao longo do ano de 2017, com ênfase no mês de maio, em que as ações do grupo tiveram seu pior desempenho desde de 2003. Foi ajustado um modelo Auto Regressivos Integrados de Médias Móveis (ARIMA) para analisar o fenômeno ocorrido, usando-se também uma análise de intervenção para obter melhores resultados. Pode-se observar que o modelo ARIMA com ponto de intervenção foi preferido em relação ao modelo ARIMA sem o ponto de intervenção, confirmando que deve-se considerar o dia 22 de maio de 2017 no modelo ajustado para a série temporal das cotações da empresa.

**Palavras-chave:** *Ibovespa, Delações Premiadas, ARIMA.*

---

<sup>1</sup>Graduanda em Economia - ICSA/UNIFAL - MG. e-mail: [raquelmaduroab@gmail.com](mailto:raquelmaduroab@gmail.com)

<sup>2</sup>Orientador - ICSA/UNIFAL - MG. e-mail: [manoel.souza@unifal-mg.edu.br](mailto:manoel.souza@unifal-mg.edu.br)

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

---

# CONTRATAÇÕES NO BRASIL SOB O REGIME DA CONSOLIDAÇÃO DAS LEIS DO TRABALHO (CLT): PERSPECTIVAS PARA O ANO DE 2018 ATRAVÉS DA ANÁLISE SÉRIES TEMPORAIS

Matheus Feres Freitas<sup>1 3</sup>

Thelma Sáfadi<sup>2 3</sup>

Haiany Aparecida Ferreira<sup>2 3</sup>

Em 1º de maio de 1943 foi sancionado por Getúlio Vargas a Consolidação das Leis do Trabalho (CLT). No Brasil, a estimativa mensal de admissões trabalhistas é feita por meio do número de trabalhadores que são contratados sob o regime da CLT. Estudos indicam que o número de admissões trabalhistas além de estar intimamente ligado à economia brasileira também tem impacto direto e indireto na educação, saúde e segurança nacional. Por este motivo é interessante estimar e/ou prever o número de contratações brasileiras, sob regime da CLT, que ocorrerão durante determinado período de tempo. Uma dentre as diversas maneiras de se obter tais previsões é através da análise e ajuste de séries temporais. Este trabalho propôs-se a ajustar e comparar modelos temporais, do tipo SARIMA, capazes de realizar previsões que estime o número de contratações, que acontecerão no Brasil, durante o ano de 2018.

**Palavras-chave:** *Admissões trabalhistas, previsões, SARIMA.*

---

<sup>1</sup> DES- UFLA. matheusferesfreitas@gmail.com

<sup>2</sup> DES- UFLA.

<sup>3</sup> Agradecimento à FAPEMIG pelo apoio financeiro.

## Determinantes da Esperança de Vida ao Nascer para o Sul/Sudoeste de Minas Gerais em 2010

Sarah de Sousa Moreira <sup>1 3</sup>

Leonardo Biazoli <sup>2 3</sup>

Marcelo Augusto Martins Teodoro <sup>2 3</sup>

Ana Carolina de Oliveira <sup>2 3</sup>

Gabriel Rodrigo Gomes Pessanha <sup>2 3</sup>

A esperança de vida ao nascer é um indicador utilizado como medida do nível de mortalidade de uma população, ou seja, se refere à quantidade de anos que se espera que um recém nascido viva. Além disso, esse indicador reflete a situação socioeconômica em que dada população se encontra. Sendo assim, este trabalho busca identificar as variáveis que influenciam a esperança de vida ao nascer do Sul/Sudoeste de Minas Gerais no ano de 2010. Com o objetivo de identificar as determinantes da esperança de vida, utilizou-se as seguintes variáveis: taxa de analfabetismo de pessoas com 15 anos ou mais, taxa de fecundidade total, taxa de emprego no setor formal e gastos municipais com saúde. Os dados referentes às variáveis foram coletados do Índice Mineiro de Responsabilidade Social (IMRS). Para a análise dos dados, aplicou-se a abordagem estatística de Regressão Linear Múltipla, empregando-a no *software* GRETL. Dessa forma, obteve-se um coeficiente de determinação ( $R^2$ ) igual a 0.5338, ou seja, as variáveis independentes presentes no modelo conseguem explicar a esperança de vida, em média, 53.38%. Essa estatística foi comprovada pela realização do teste  $F$  que se mostrou significativo ao nível de 1%. Os pressupostos teóricos da regressão linear múltipla (normalidade dos resíduos, ausência de colinearidade, ausência de heterocedasticidade, ausência de autocorrelação dos resíduos e independência dos parâmetros) foram testados e respeitados. A taxa de analfabetismo e a taxa de fecundidade representam uma relação inversa com a esperança de vida, enquanto a taxa de emprego no setor formal e os gastos com saúde indicam uma relação direta.

**Palavras-chave:** *Esperança de Vida ao Nascer, Regressão Linear Múltipla, Minas Gerais.*

---

<sup>1</sup>ICSA - UNIFAL-MG. e-mail: [sarah.souza63@gmail.com](mailto:sarah.souza63@gmail.com)

<sup>2</sup>ICSA - UNIFAL-MG.

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## Estudo dos modelos de precificação de ativos em uma instituição bancária

Poliana Maria Benelli <sup>1 3</sup>

Gabriel Rodrigo Gomes Pessanha <sup>2 3</sup>

Modelos de precificação de ativos são utilizados em finanças para determinar a taxa de retorno apropriada de um determinado ativo em relação a uma carteira de mercado perfeitamente diversificada. São utilizados também para estimar o custo do capital próprio de uma empresa. O *Capital Asset Pricing Model* (CAPM) é o modelo de precificação mais conhecido. Sendo o CAPM considerado por muitos autores simplista e irreal, surgiram outros modelos de precificação, como é o caso do *Arbitrage Pricing Theory* (APT). A notável diferença entre os modelos a quantidade de fatores relevantes de risco. O objetivo deste trabalho é comparar os dois modelos de precificação de ativos, CAPM e APT, para averiguar qual deles é mais preciso na precificação de um ativo do setor bancário a partir de 1995. A amostra é composta por ações de um dos três maiores bancos em valor de ativos listados na Bovespa: Bradesco. Os dados utilizados são trimestrais e compreendidos entre o terceiro trimestre de 1997 e o terceiro trimestre de 2017. Os valores coletados e tratados foram importados para os *softwares* gratuitos Gretl e R, onde foram realizadas as regressões lineares simples (RLS) e múltiplas (RLM) via método de Mínimos Quadrados Ordinários (MQO). O coeficiente de determinação  $R^2$  do modelo APT mostrou-se superior em todas regressões para ambos os modelos.

Palavras-chave: APT; CAPM; Precificação de ativos.

---

<sup>1</sup> ICSA - UNIFAL. e-mail polianabenelli@gmail.com

<sup>2</sup> ICSA – UNIFAL.

<sup>3</sup> Agradecimento à FAPEMIG pelo apoio financeiro.

## **Estudo econométrico para o índice de confiança do consumidor em função de variáveis macroeconômicas no período de 2005 a 2015**

**Bruna Leandra Zeferino Silva**<sup>1 3</sup>

**Letícia Lima Milani**<sup>2 3</sup>

O Índice de Confiança do Consumidor (ICC), divulgado pelo Instituto Brasileiro de Economia (IBRE/FGV), tem como objetivo mensurar as expectativas dos consumidores no que se refere à situação econômica atual e futura do país, analisando suas perspectivas de consumo, poupança, mercado de trabalho, entre outros. Portanto, o intuito deste trabalho foi verificar se o ICC, considerado no período de setembro de 2005 a dezembro de 2015, pode ser explicado pelas variáveis macroeconômicas: Taxa de desocupação, Risco Brasil, Taxa de juros Selic e Poupança, que influenciam economicamente o comportamento do mercado, tanto de forma positiva quanto negativa, afetando, principalmente, os hábitos de consumo e rendimento da população. Para esta análise foi utilizada a regressão linear múltipla ajustada pelo método de mínimos quadrados ordinários, apresentando resultados estatisticamente significantes a um grau de significância de 1%. O modelo estudado demonstrou através do  $R^2$  que as variações do índice de confiança do consumidor são explicadas pelos determinantes econômicos em 86,72 % dos casos, satisfazendo os testes de normalidade dos resíduos, homocedasticidade, colinearidade e especificação do modelo.

**Palavras-chave:** Índice de Confiança do Consumidor, Variáveis Macroeconômicas, Regressão Linear Múltipla.

---

<sup>1</sup> ICSA – UNIFAL-MG, e-mail: bru.leandra27@hotmail.com

<sup>2</sup> Professora Adjunta UNIFAL-MG campus Varginha

<sup>3</sup> Agradecimento ao CNPq e FAPEMIG pelo apoio financeiro

# Eficiência de mercado do índice Ibovespa usando DFA

Kerolly Kedma Felix do Nascimento <sup>1 3</sup>

Fabio Sandro dos Santos <sup>2</sup>

Elielma Santana de Jesus <sup>2 3</sup>

Jader da Silva Jale <sup>2 3</sup>

Tiago Alessandro Espínola Ferreira <sup>2</sup>

Neste trabalho, analisamos a Eficiência de Mercado (EM) para o índice Ibovespa, estimando o expoente de Hurst ( $H$ ) por meio da análise de flutuação sem tendência (DFA-*Detrended Fluctuation Analysis*). Calculamos o expoente de Hurst em janelas móveis (4000 subamostras) de um ano (240 dias úteis) para o log-retorno do índice Ibovespa entre 01/01/2001 e 31/12/2017. Nossos resultados indicam que  $\sim 41\%$  das subamostras apresentam comportamento de longo alcance persistente ( $H > 0,5$ ), ou seja, maiores/menores valores têm maior probabilidade de serem seguidos por maiores/menores valores. Em contrapartida, o comportamento de longo alcance antipersistente ( $H < 0,5$ ) é observado em  $\sim 59\%$  das subamostras, em que maiores/menores valores têm maior probabilidade de serem seguidos por menores/menores valores. Realizamos testes t-Student, à 5% de significância, sobre a série temporal do expoente de Hurst  $H$ , usando janelas móveis de tamanho 30 (3971 subamostras) para a hipótese de nulidade  $H_0 : H = 0,5$  contra a hipótese alternativa  $H_0 : H \neq 0,5$ . Encontramos que a Eficiência de Mercado ( $H_0 : H = 0,5$ ) foi aceita em apenas  $\sim 19\%$  das subamostras. Este resultado indica que o índice Ibovespa não se comporta de acordo com a Hipótese de Eficiência de Mercado (HEM).

**Palavras-chave:** *DFA, Eficiência de mercado, Hurst, Ibovespa.*

---

<sup>1</sup>DEINFO - UFRPE. e-mail: [kerollyfn@gmail.com](mailto:kerollyfn@gmail.com)

<sup>2</sup>DEINFO - UFRPE.

<sup>3</sup>Agradecimento à CAPES pelo apoio financeiro.

## Fluxo de informação entre Ibovespa e taxa de câmbio na crise *subprime*

Kerolly Kedma Felix do Nascimento <sup>1 3</sup>

Fabio Sandro dos Santos <sup>2</sup>

Josafá José do Carmo Reis Júnior <sup>2 3</sup>

Jader da Silva Jale <sup>2 3</sup>

Tiago Alessandro Espínola Ferreira <sup>2</sup>

Neste trabalho, estudamos a direção do fluxo de informação entre o índice Ibovespa e a taxa de câmbio usando o método Transfer Entropy (TE), diante da crise *Subprime*. Analisamos três cenários distintos: períodos pré-crise (01/01/2001 – 31/07/2007), crise (01/08/2007 – 31/03/2009) e pós-crise financeira mundial (01/04/2009 – 31/12/2017). No período anterior a crise obtivemos  $TE(\text{Ibovespa} \rightarrow \text{Câmbio}) = 0,0624$  contra  $TE(\text{Câmbio} \rightarrow \text{Ibovespa}) = 0,0720$ . Na crise,  $TE(\text{Ibovespa} \rightarrow \text{Câmbio}) = 0,0140$  contra  $TE(\text{Câmbio} \rightarrow \text{Ibovespa}) = 0,1136$ . Finalmente, para o período pós-crise,  $TE(\text{Ibovespa} \rightarrow \text{Câmbio}) = 0,0517$  e  $TE(\text{Câmbio} \rightarrow \text{Ibovespa}) = 0,0621$ . Nossos resultados indicam que no período da crise (01/08/2007 – 31/03/2009), o fluxo de informação da taxa de câmbio para o índice Ibovespa foi  $(0.1136/0.0140) \sim 8$  vezes maior que no sentido contrário. Este resultado confirma que o método usado é promissor para quantificar o fluxo de informação entre duas séries temporais simultâneas. Além disso, os resultados obtidos refletem a agitação e incerteza no mercado brasileiro durante a pior crise financeira do capitalismo moderno, desde a grande crise de 1929.

**Palavras-chave:** *Crise Subprime, Ibovespa, Taxa de Câmbio, Transfer Entropy.*

---

<sup>1</sup>DEINFO - UFRPE. e-mail: [kerollyfn@gmail.com](mailto:kerollyfn@gmail.com)

<sup>2</sup>DEINFO - UFRPE.

<sup>3</sup>Agradecimento à CAPES pelo apoio financeiro.



## Função Resposta Impulso aplicadas às séries econômicas: Produção Industrial e Índice de Incerteza da Economia

Aline Caroline Rodrigues <sup>1 3</sup>

Letícia Lima Milani <sup>2 3</sup>

A função resposta impulso demonstra como um choque em qualquer das variáveis afeta todas as demais variáveis endógenas, e, eventualmente reflete sobre a própria variável (Hiron; Telma, 2008). Em se tratando de economia, essa ferramenta permite verificar como as variáveis endógenas reagirão, podendo auferir possíveis alterações em suas séries, dado um período de tempo. Desta forma, objetiva-se, analisar a função resposta impulso nas séries produção industrial (PI) e índice de incerteza da econômica (IIE-Br), ambas obtidas no IPEADATA, sendo sua abrangência mensal de 2006 até 2016. As análises foram realizadas por meio do *software* estatístico *R*. O primeiro passo para análise foi a construção de um modelo vetorial autorregressivo (VAR), para seleção de defasagens utilizou o critério de informação de Akaike (AIC), o qual sugeriu três defasagens. Após a construção do VAR, gerou-se a função impulso resposta e seus gráficos. Como resultados pode-se observar que um choque na variável PI gera uma queda abrupta em IIE-Br nos dois primeiros meses, após esse período ocorre uma leve recuperação que é perdida no quinto mês, a partir desta data sofre pequenas oscilações e se estabiliza. Em relação ao choque em IIE-Br, constata-se uma queda constante até o segundo mês, em seguida uma queda mais amena até o quinto mês, após pequenas oscilações o choque também se estabiliza. Por meio deste ferramental foi perceptível que em ambas as variáveis o choque afeta de forma negativa até o quinto mês, após esse período o choque se perde e ocorre a estabilização.

**Palavras-chave:** *Econometria, Impulso Resposta, VAR.*

---

<sup>1</sup>ICSA - Unifal-MG e-mail: [alineacr25@gmail.com](mailto:alineacr25@gmail.com)

<sup>2</sup>ICSA - Unifal-MG.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

## Função Resposta Impulso aplicadas às séries econômicas: Produção Industrial e Índice de Incerteza da Economia

Aline Caroline Rodrigues <sup>1 3</sup>

Letícia Lima Milani <sup>2 3</sup>

A função resposta impulso demonstra como um choque em qualquer das variáveis afeta todas as demais variáveis endógenas, e, eventualmente reflete sobre a própria variável (Hiron; Telma, 2008). Em se tratando de economia, essa ferramenta permite verificar como as variáveis endógenas reagirão, podendo auferir possíveis alterações em suas séries, dado um período de tempo. Desta forma, objetiva-se, analisar a função resposta impulso nas séries produção industrial (PI) e índice de incerteza da econômica (IIE-Br), ambas obtidas no IPEADATA, sendo sua abrangência mensal de 2006 até 2016. As análises foram realizadas por meio do *software* estatístico *R*. O primeiro passo para análise foi a construção de um modelo vetorial autorregressivo (VAR), para seleção de defasagens utilizou o critério de informação de Akaike (AIC), o qual sugeriu três defasagens. Após a construção do VAR, gerou-se a função impulso resposta e seus gráficos. Como resultados pode-se observar que um choque na variável PI gera uma queda abrupta em IIE-Br nos dois primeiros meses, após esse período ocorre uma leve recuperação que é perdida no quinto mês, a partir desta data sofre pequenas oscilações e se estabiliza. Em relação ao choque em IIE-Br, constata-se uma queda constante até o segundo mês, em seguida uma queda mais amena até o quinto mês, após pequenas oscilações o choque também se estabiliza. Por meio deste ferramental foi perceptível que em ambas as variáveis o choque afeta de forma negativa até o quinto mês, após esse período o choque se perde e ocorre a estabilização.

**Palavras-chave:** *Econometria, Impulso Resposta, VAR.*

---

<sup>1</sup>ICSA - Unifal-MG e-mail: [alineacr25@gmail.com](mailto:alineacr25@gmail.com)

<sup>2</sup>ICSA - Unifal-MG.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

## Identificação de clusters a partir de método não-hierárquico: uma análise socioeconômica do Estado de Minas Gerais

Leonardo Henrique de Andrade B. do Nascimento<sup>1 3</sup>

Ana Carolina de Oliveira<sup>2 3</sup>

Este trabalho consiste em agrupar os municípios do Estado de Minas Gerais (MG) utilizando a técnica multivariada de análise de agrupamento. O objetivo é identificar os grupos de municípios mais desenvolvidos com base em variáveis socioeconômicas através do 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 853 municípios presentes em MG e foram analisados a partir de 10 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 grupos, sendo eles: grupo 1 ( $G1$ ) - 297 municípios; grupo 2 ( $G2$ ) - 128 municípios; grupo 3 ( $G3$ ) - 309 municípios; e o grupo 4 ( $G4$ ) - 119 municípios. 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. Observado o mapa da região, construído com relação aos grupos, percebe-se que há uma concentração dos municípios alocados no  $G3$  e  $G4$  em uma região específica, entretanto não é possível comprovar uma relação espacial a partir da metodologia adotada no estudo. Após a análise de agrupamento, sugere-se que o  $G4$  possui um desenvolvimento socioeconômico superior visto que dentre as 70 maiores cidades do Estado de MG, 59 estão presentes nesse grupo.

**Palavras-chave:** *Minas Gerais, Clusters, Análise Multivariada.*

---

<sup>1</sup>ICSA - UNIFAL-MG. e-mail: [leonardoandrade.h@outlook.com](mailto:leonardoandrade.h@outlook.com)

<sup>2</sup>ICSA - UNIFAL-MG. e-mail: [ana.coliveira@outlook.com](mailto:ana.coliveira@outlook.com)

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## Influência da previdência fechada e do envelhecimento populacional na previdência aberta

Gabriel Alves São Severino<sup>1 4</sup>

Josiane Correia de Souza Carvalho<sup>2 3 4</sup>

Fabrizio Avelar<sup>2 4</sup>

O sistema previdenciário brasileiro é dividido em previdência social (obrigatória), e previdência complementar (facultativa). A previdência complementar possui duas subdivisões: a previdência fechada, instituída por uma empresa e acessível a quem possua vínculo empregatício, e a aberta, na qual qualquer pessoa pode adquirir um plano de benefícios. Com as propostas de reforma na previdência social, justificada pelo envelhecimento populacional, discussões acerca da previdência aberta tornam-se cada vez mais importantes, por ser uma outra alternativa disponível a todos. Neste estudo busca-se avaliar o comportamento da variável previdência aberta a partir do comportamento da previdência fechada e envelhecimento populacional. Para isso o modelo de regressão linear múltiplo foi ajustado aos dados de previdência, considerando-se como variável dependente a arrecadação com previdência aberta e, como variáveis independentes, a arrecadação com previdência fechada e o índice de envelhecimento da população brasileira. As duas variáveis independentes foram significativas, ao nível de 5% de significância. O teste de Shapiro-Wilks foi utilizado para se verificar a normalidade dos resíduos e a homogeneidade da variância foi examinada por meio do teste de Breusch-Pagan. Ao nível de 5% de significância, os resíduos foram considerados normalmente distribuídos e com variância constante. Desta forma, os resultados demonstram que os ramos da previdência complementar (aberta e fechada), crescem simultaneamente, o que pode indicar que as pessoas optam por realizar mais que uma previdência. Já a associação do envelhecimento populacional com a previdência complementar aberta, demonstra que as pessoas tendem a procurar outra forma de garantir estabilidade no futuro, dado o aumento neste fator que é importante para o equilíbrio da previdência social.

**Palavras-chave:** *Regressão linear múltipla, Previdência Complementar, Previdência Social.*

---

<sup>1</sup>Discente de Ciências Atuariais- UNIFAL-MG. e-mail: [gabrielalvesan@gmail.com](mailto:gabrielalvesan@gmail.com)

<sup>2</sup>Programa de Pós Graduação em Estatística Aplicada e Biometria- UNIFAL-MG.

<sup>3</sup>Agradecimento à CAPES pela bolsa.

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## Métodos Multivariados na análise do Sul/Sudoeste de Minas Gerais: Uma aplicação com variáveis Socioeconômicas e número de beneficiários de Planos de Saúde

Leonardo Biazoli <sup>1 3</sup>

Sarah de Sousa Moreira <sup>2 3</sup>

Ana Carolina de Oliveira <sup>2 3</sup>

Leonardo Henrique de Andrade B. do Nascimento <sup>2 3</sup>

Este trabalho tem como objetivo agrupar os municípios do Sul/Sudoeste de Minas Gerais que apresentam características similares ao relacionar as seguintes variáveis: número de beneficiários de Operadoras de Planos de Saúde (OPS), taxa de fecundidade, esperança de vida, renda per capita e taxa de emprego no setor formal. O número de beneficiário de OPS foi extraído do banco de dados da Agência Nacional de Saúde Suplementar (ANS), e os demais foram retirados do Índice Mineiro de Responsabilidade Social (IMRS). Utilizou-se a técnica multivariada de análise de agrupamento. Aplicou-se o método hierárquico de Ward, com o intuito de definir o número de grupos a partir do ponto de corte com maior altura livre do dendrograma. Definido o número de grupos, aplicou-se o método não-hierárquico das *k*-médias, alocando os municípios em quatro grupos, sendo eles: grupo 1 (*G1*) - 117 municípios, grupo 2 (*G2*) - 4 municípios, grupo 3 (*G3*) - 17 municípios e grupo 4 (*G4*) - 8 municípios. Quando calculada a mediana das variáveis, o *G2* apresentou os valores mais altos para as variáveis, sendo eles: taxa de fecundidade (1.81), esperança de vida (78.06), beneficiários (63938), renda *per capita* (R\$953.96) e taxa de emprego formal (44.7). Através do coeficiente de variação, pode-se observar que o *G2* teve uma menor variação realtiva dos dados. Observa-se que dos 5 municípios que apresentam as maiores populações da mesorregião, 4 deles estão no *G4*. Além disso, nesses municípios existem OPS, o que pode justificar o comportamento dos dados ao serem agrupados.

**Palavras-chave:** *Análise de Agrupamento, Análise Multivariada, Operadoras de Plano de Saúde.*

<sup>1</sup>Universidade Federal de Alfenas - UNIFAL-MG. e-mail: [leonardobiazoli19@gmail.com](mailto:leonardobiazoli19@gmail.com)

<sup>2</sup>Universidade Federal de Alfenas - UNIFAL-MG.

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## **Modelagem da Poupança Através da Perspectiva Econômica: Uma aplicação do MRLM**

**Carlos Ayallas de Sousa Pereira**<sup>1 3</sup>

**Zeferino Gomes da Silva Neto**<sup>1 3</sup>

**Valmária Rocha da Silva Ferraz**<sup>2</sup>

**Diogo Bona Ibiapina**<sup>2</sup>

A caderneta de poupança foi proposta em 1861 com a missão inicial de atender as classes mais populares. A quantia gerada nos depósitos é de suma importância para o subsídio de programas governamentais. Tendo em vista a importância do mais antigo e popular investimento brasileiro, este trabalho propõe avaliar os valores depositados na caderneta de poupança verificando a influência do Produto Interno Bruto (PIB) e do Índice Nacional de Preços ao Consumidor Amplo (IPCA). Para tal utilizou-se o método de Regressão Linear Múltipla (RLM), propondo um modelo para previsões. Foram usados dados distribuídos mensalmente, no período de março de 2012 à março de 2017. Por meio da ANOVA constatou-se que as variáveis foram significativas para o modelo. A presença de valores influentes nos dados iniciais obrigou a remoção de tais para que o modelo pudesse ser ajustado corretamente. Após a remoção os pressupostos do RLM não foram violados. O modelo proposto explica 93,1% da variação dos depósitos na poupança, e indica a sua adequabilidade para a realização de inferências.

**Palavras-chave:** *Inflação, Poupança, Regressão Linear Múltipla.*

---

<sup>1</sup> DET – UFV. e-mail carlosayallas@hotmail.com

<sup>2</sup> UFPI

<sup>3</sup> Agradecimento à FAPEMIG e a CAPES pelo apoio financeiro.

# Modelagem do número de certificações ISO 9001 emitidos para o Brasil e Chile

Luiz Henrique Marra da Silva Ribeiro <sup>1 5</sup>

Luiz Alberto Beijo <sup>2 5</sup>

Denismar Alves Nogueira <sup>3 5</sup>

Eduardo Gomes Salgado <sup>4 5</sup>

A ISO 9001 é considerada uma das mais efetivas ferramentas suporte para o sistema de gerenciamento da qualidade, pela qual, gerentes e pesquisadores têm interesse na evolução e fatores que influenciam o número de certificações nos países. Globalmente, o número de certificações dessa norma vem aumentando consideravelmente devido a demandas de setores públicos e privados. Porém, estudos sobre a implementação e difusão da ISO 9001 ainda são escassos. Diante disso, o objetivo desse trabalho é verificar quais fatores influenciam o número de ISO 9001 no Brasil e Chile, no período de 2003 à 2016, ajudando na compreensão de quais fatores estão influenciando o número de ISO 9001 nesses países. Foram analisados: produto interno bruto (PIB), produto nacional bruto (PNB), porcentagem de exportação e reservas totais, em valores anuais, também foram utilizados valores de um e dois anos anteriores, e valores per capita. Foram ajustados modelos de regressão linear, sendo que os resíduos apresentaram: normalidade pelo teste de Shapiro-Wilk, homoscedasticidade pelo teste de Breusch-Pagan, independência pelo teste de Durbin-Watson. Para o Brasil, foram significativas ao nível de 5%: reservas totais com efeito negativo e produto interno bruto com efeito positivo, apresentando  $R^2 = 84,29\%$ . Para o Chile, a porcentagem de exportação, PNB de dois anos anteriores e número de certificações ISO 9001 de dois anos anteriores tiveram efeito positivo, enquanto as reservas totais tiveram efeito negativo, todas significativas ao nível de 5%, apresentando  $R^2 = 95,46\%$ .

**Palavras-chave:** *Gerenciamento da qualidade, Produto interno bruto, Regressão linear, Reservas totais, Porcentagem de exportação.*

---

<sup>1</sup>Mestrando no PPGEAB - Unifal-MG. e-mail: [luiz.marra@outlook.com](mailto:luiz.marra@outlook.com).

<sup>2</sup>Professor Doutor no PPGEAB - Unifal-MG. e-mail: [luiz.beijo@unifal-mg.edu.br](mailto:luiz.beijo@unifal-mg.edu.br).

<sup>3</sup>Professor Doutor no PPGEAB - Unifal-MG. e-mail: [denismar.nogueira@unifal-mg.edu.br](mailto:denismar.nogueira@unifal-mg.edu.br).

<sup>4</sup>Professor Doutor no Departamento de Ciência da Computação - Unifal-MG.  
e-mail: [eduardosalgado@bcc.unifal-mg.edu.br](mailto:eduardosalgado@bcc.unifal-mg.edu.br).

<sup>5</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## Modelo de Séries Temporais Aplicado à Série da Intenção de Consumo das Famílias Brasileiras

Camila Moreira Ribeiro<sup>1 3</sup>

Letícia Lima Milani<sup>2 3</sup>

**Resumo:** A Pesquisa de Intenção de Consumo das Famílias (ICF), realizada pela Confederação Nacional do Comércio de Bens, Serviços e Turismo (CNC), é um indicador que tem como objetivo medir a avaliação que os consumidores fazem sobre aspectos importantes da condição de vida de sua família, tais como a sua capacidade de consumo (atual e de curto prazo), nível de renda doméstico, segurança no emprego e qualidade de consumo, presente e futuro. Em outras palavras, segundo o CNC, o ICF é um indicador antecedente do consumo, a partir do ponto de vista dos consumidores, tornando-o uma ferramenta poderosa para o planejamento do comércio e de outras atividades produtivas. Dessa forma, diante do exposto, o objetivo é ajustar um modelo de Box e Jenkins para se fazer previsões do índice da Intenção de Consumo das Famílias para os meses de outubro, novembro e dezembro de 2017. Para que os modelos pudessem ser identificados e selecionados, analisou-se o comportamento da série, verificando efeitos de tendência e sazonalidade. Após a retirada das componentes tendência e sazonalidade por meio de diferenciações, identificou-se um modelo  $SARIMA(p,d,q)(P,D,Q)_{12}$ . Dentre alguns modelos, o que apresentou menor valor de AIC foi selecionado e escolheu-se o  $SARIMA(1,1,1)(0,1,1)_{12}$  para que a previsão pudesse ser realizada.

**Palavras-chave:** *Box-Jenkins, Consumo das Famílias, SARIMA.*

---

<sup>1</sup> ICSA – UNIFAL, email: camilam.ribeiro@yahoo.com.br

<sup>2</sup> Professora Adjunta UNIFAL-MG *campus* Varginha.

<sup>3</sup> Agradecimento a FAPEMIG pelo apoio financeiro.



# Modelos de Machine Learning na Predição do Preço do Café

Lucas Pereira Lopes <sup>1</sup>

O conhecimento do comportamento dos preços torna-se extremamente útil nas tomadas de decisões por parte dos produtores com relação ao planejamento da produção e a formação de estoques. Vários trabalhos na literatura fazem a predição do preço de commodities agrícolas com o auxílio de variáveis explicativas utilizando métodos de regressão via mínimos quadrados. Geralmente, esses modelos lineares padrão assumem que os preditores têm uma exogeneidade fraca, linearidade, homocedasticidade, distribuição dos resíduos normal e falta de multicolinearidade. No entanto, trabalhando com variáveis dependentes e independentes que sejam de natureza econômicas, muitas dessas suposições são falhas. Uma boa abordagem para fazer predições são métodos conhecidos como *statistical machine learning*. Portanto, o objetivo deste trabalho é prever o preço do café brasileiro e, além disso, analisar o comportamento de variáveis explicativas que podem influenciar o mercado de café, possibilitando um ganho por parte dos produtores em relação a maximização do lucro e minimização de riscos. De acordo com trabalhos correlacionados já realizados na literatura, as variáveis explicativas utilizadas neste trabalho são: taxa de câmbio, taxa de juros, crédito rural, PIB do Brasil e dos principais importadores de café brasileiro e preços dos principais cafés produzidos pelos dois principais concorrentes em termo de produção (vietnã e colômbia). As métricas utilizadas para medir a qualidade preditiva dos modelos são: *Mean Absolute Percentage Error* (MAPE), *Root Mean Squared Error* (RMSE) e *Mean Absolute Error* (MAE). Como resultado, obteve-se a concordância entre os critérios de qualidade de ajuste e por isso será mencionado o critério MAPE. O melhor modelo em relação ao poder preditivo fora da amostra foi o *Support Vector Machine* com Kernel Linear, mostrando uma diferença percentual entre o predito e o observado de 2,99%, seguido dos modelos LASSO (3,78%), SVM com Kernel Gaussiano (5,08%), Boosting (5,36%), Árvore de Regressão (6,8%), K-NN (8,96%), Floretas Aleatórias (11,46%) e ARIMA (18,8%). Este resultado mostra o alto poder preditivo da técnica SVM com kernel linear. Além da obtenção do alto poder preditivo dos modelos supracitados, é possível extrair quais variáveis que impactam e, em qual magnitude, o preço do café brasileiro, os quais serão discutidos no evento. Portanto, a elaboração e utilização de modelos com alta acurácia no meio agrícola possibilita utilizar das estratégias das fases de baixa e de alta nos preços para criar um ambiente de maximização dos lucros.

**Palavras-chave:** *Café, Machine Learning, Preço, Predição.*

---

<sup>1</sup>ICMC - USP/UFSCar. e-mail: [lucas.lopes@usp.br](mailto:lucas.lopes@usp.br)

## Modelos de previsão para o crédito bancário em Sergipe

William das Virgens Souza <sup>1</sup>

Marcos Santos Oliveira <sup>2</sup>

O crédito é o principal fator a movimentar uma empresa e desenvolvê-la, sendo uma importante fonte de financiamento em situações de descasamento de caixa ou de necessidade de investimentos para modernização e manutenção da capacidade produtiva. A partir de 2003, a economia brasileira passou por uma forte expansão do crédito, no qual as instituições financeiras tiveram um papel significativo nesse processo com as chamadas operações de crédito, aumentando seu grau de alavancagem e propensão ao risco. O crédito ao consumidor proporciona o crescimento da economia, eleva tanto a capacidade de compra, aumentando o número de transações comerciais e atuando como um redutor de restrição orçamentária. Dessa forma, o objetivo deste estudo busca aplicar a técnica de análise de séries temporais com aplicação da metodologia de Box & Jenkins, para os saldos das operações de créditos do estado de Sergipe, tendo em vista identificar um modelo econométrico com realização de previsão de períodos futuros. O modelo SARIMA(2,1,1)(1,0,2)[12] apresentou o melhor ajuste aos dados, haja vista que a escolha do modelo se deu ao fato do mesmo possuir o menor Erro percentual absoluto médio (MAPE), 0,44%.

**Palavras-chave:** *Metodologia Box & Jenkins. Sarima. Crédito.*

---

<sup>1</sup> DECAT – UFS. e-mail william.cbf21@gmail.com

<sup>2</sup> DECAT – UFS. e-mail marcos@exogenia.com.br

---

## MODELOS DE SAZONALIDADE APLICADOS PARA SÉRIES DE SUPRIMENTO DE FUNDOS DA UNIVERSIDADE FEDERAL DE SANTA MARIA

Vanessa Siqueira Peres da Silva<sup>1 4</sup>

Michael Gonçalves da Silva<sup>2 4</sup>

Fernanda Dalcin Flôres<sup>3 4</sup>

### *Resumo*

O presente trabalho tem por objetivo fazer uma análise do comportamento das séries financeira dos suprimentos de fundos da Universidade Federal de Santa Maria - UFSM, de janeiro de 2013 até julho de 2017. Utilizando-se a metodologia de Box & Jenkins seguiu-se com a identificação, estimação e verificação do modelo e posterior previsão. Desta forma, com a análise do correlograma e periodograma foi possível encontrar o componente sazonal na série. Foram escolhidos modelos da classe ARIMA com componente sazonal (SARIMA), utilizando os critérios de AIC e HCQ. Selecionando aqueles que possuísem valores mais baixos, o modelo escolhido foi o SARIMA (1,1,1)(1,1,0). Após escolher o modelo, o mesmo foi estimado obtendo-se os valores dos coeficientes, e em seguida fez-se a análise dos resíduos do modelo para verificar e confirmar a existência de ruído branco. Com os valores estimados realizou-se a previsão para vinte quatro meses, e constatou-se a presença de sazonalidade, tendo uma maior utilização deste instrumento nos últimos três meses finais do exercício financeiro.

***Palavras-chave:*** Box & Jenkins. Suprimento de Fundos. Modelo SARIMA.

---

<sup>1</sup> CCNE - UFSM. e-mail [vanessa@ufsm.br](mailto:vanessa@ufsm.br)

<sup>2</sup> UNIFRA.

<sup>3</sup> CCSH – UFSM.

<sup>4</sup> Agradecimento ao CNPq pelo apoio financeiro.

---

**O valor do ensino médio:  
o impacto da escolaridade sobre o rendimento e a inserção no mercado de trabalho**

**Leda Grasielle Oliveira<sup>1</sup>**

**Lincoln Frias<sup>2</sup>**

Sessenta por cento dos brasileiros maiores de 25 anos não concluíram o ensino médio (IBGE-PNAD, 2017). Esse fato provavelmente é um dos principais obstáculos para o desenvolvimento social e econômico do país. O objetivo deste trabalho é avaliar o efeito do ensino médio sobre a inserção dos indivíduos no mercado de trabalho, comparando a situação das pessoas cujo nível máximo de instrução é o ensino médio com aquelas que concluíram apenas o ensino fundamental, analisando os rendimentos e os principais tipos de ocupação. Os dados são referentes ao segundo trimestre de 2017 e foram retirados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC) do IBGE. Utilizando a linguagem *Python*, por meio da interface *Jupyter*, além da análise descritiva, foi realizada a análise de regressão múltipla, enfatizando-se sua interpretação causal (ANGRIST; PISCHKE, 2009). Foram construídos seis modelos, empregando erros padrão robustos, usando diferentes covariáveis e diferentes restrições da amostra. Todos os modelos indicaram que o ensino médio causa um acréscimo no rendimento médio. O valor deste acréscimo variou de acordo com o modelo, indo de R\$205,50 a R\$390,50, mas sempre significativo ao nível de significância de 5%. Além disso, o efeito do ensino médio é sensível ao sexo, as mulheres tendem a ter rendimentos de R\$491,00 a R\$625,00 menores. Por fim, enquanto as ocupações típicas do ensino fundamental são os trabalhadores domésticos (8%), trabalhadores agrícolas (6%) e pedreiros (5%), no caso do ensino médio predominam balconistas e vendedores (7%), trabalhadores domésticos (4%) e assistentes administrativos (4%).

**Palavras-chave:** *Inferência Causal, Mercado de Trabalho, Regressão Linear.*

---

1 PPGPS-Unifal (Prog. Pós-Grad. Gestão Pública e Sociedade - Universidade Federal de Alfenas, campus Varginha), ledagrasiele@gmail.com. Agradecimento à Unifal pela bolsa de mestrado.

2 PPGPS-Unifal, campus Varginha, lincolnfrias@gmail.com.

# Predição e Determinantes de Crimes de Homicídios no Brasil utilizando Métodos de Machine Learning não Paramétricos

Lucas Pereira Lopes <sup>1</sup>

Beatriz Rezzieri Marchezini <sup>2</sup>

A literatura traz vários trabalhos sobre a predição e determinantes de crimes com o auxílio de variáveis explicativas utilizando métodos de regressão via mínimos quadrados. Geralmente, esses modelos lineares padrão assumem que os preditores têm uma exogeneidade fraca (variáveis livres de erro), linearidade, homocedasticidade, distribuição dos resíduos normal e falta de multicolinearidade. Logo, quando essas suposições não estão satisfeitas, as conclusões sobre os determinantes estão suscetíveis a considerações equívocas e divergentes. Em contrapartida, os modelos de regressão não paramétricos da classe de aprendizado de máquina estatístico não fazem as suposições supracitadas. Assim, essas técnicas não paramétricas não fazem hipóteses sobre o relacionamento das covariáveis com a variável resposta, ou seja, são ferramentas que não possuem pressuposições sobre as distribuições das variáveis e não afirmam sobre alguma forma funcional entre as mesmas. Portanto, este trabalho tem dois principais objetivos: 1) utilizar técnicas de aprendizado de máquina estatístico para prever crimes de homicídios nos 5.570 municípios brasileiros, onde os métodos são: Árvore de Regressão, Florestas Aleatórias, Boosting e K-Nearest Neighbors e 2) além da busca pelo melhor algoritmo para predição, determinar quais variáveis que mais interferem na ocorrência de crimes de homicídios em nível nacional. Como resultado, notou-se a alta capacidade de predição do modelo de florestas aleatórias (explicando 85% da variabilidade total e um  $R^2$  de, em média, 90%). Além disso, os métodos corroboraram entre si em suas predições e, pode-se extrair do melhor modelo ajustado quais variáveis interferem em nível nacional o número de homicídios.

**Palavras-chave:** *Crime, Determinantes, Homicídios, Machine Learning, Predição.*

---

<sup>1</sup>ICMC - USP/UFSCar. e-mail: [lucas.lope@usp.br](mailto:lucas.lope@usp.br)

<sup>2</sup>UFABC. e-mail: [beatriz.marchezini@ufabc.edu.br](mailto:beatriz.marchezini@ufabc.edu.br)

## Previsão do retorno anormal acumulado no setor bancário utilizando redes neurais artificiais

Matheus Saraiva Alcino <sup>1 3</sup>

Gabriel Rodrigo Gomes Pessanha <sup>2 3</sup>

O mercado acionário é uma rede que interliga setores dos mais diversos ramos e, com isso, cada decisão tomada pode ocasionar uma série de desdobramentos, que por sua vez afetam direta ou indiretamente o retorno de investidores. Por causa disso, entender o comportamento ou simplesmente conhecer as tendências do mercado pode ajudar na tomada de decisão. Assim, o objetivo deste trabalho é analisar o retorno anormal acumulado no setor bancário a fim de se atingir uma previsão significativa deste retorno. Para isso, utilizou-se dados referentes aos cinco maiores bancos brasileiros de capital aberto na bolsa de valores BM&FBovespa. Estes dados foram retirados do programa Economática e análise consiste no método estatístico de redes neurais artificiais. Os resultados mostram que a rede atingiu um alto poder de previsão a partir dos determinantes utilizados, e com isso, acredita-se que as redes neurais artificiais podem ajudar na criação de melhores estratégias de investimento.

**Palavras-chave:** *Retorno anormal acumulado, redes neurais artificiais, mercado acionário, previsão, setor bancário.*

---

<sup>1</sup>ICSA - UNIFAL. e-mail: [osaraivamatheus@gmail.com](mailto:osaraivamatheus@gmail.com)

<sup>2</sup>ICSA - UNIFAL. e-mail: [gabrielrgp@yahoo.com.br](mailto:gabrielrgp@yahoo.com.br)

<sup>3</sup>Agradecimento a Fapemig pelo apoio financeiro.

## Proposta de modelagem de risco atuarial para proporcionar rentabilidade

Geniffer Sabrina Brito da Hora<sup>1</sup>

Marcos Santos Oliveira<sup>2</sup>

O presente trabalho teve como objetivo principal analisar o modelo de *Credit Scoring*, a fim de verificar a capacidade discriminatória do modelo em uma carteira de convênio odontológico e observar em qual perfil de cliente a operadora poderá obter rentabilidade. Neste estudo foram demonstradas as causas e necessidades que levam a população a adquirir planos odontológicos, além de demonstrar os riscos que a instituição se expõe ao prestar determinado serviço. As informações utilizadas nesta pesquisa foram disponibilizadas pela operadora e com estes dados foram criadas variáveis *dummies* que foram obtidas através do Risco Relativo e da matriz de correlação de Spearman. Na análise destes dados foi utilizado o modelo de Regressão Logística Binária, desenvolvido através do software R-project, o modelo foi validado com o teste de Kolmogorov-Smirnov (KS), curva ROC e Hosmer-Lemeshow, após validação do modelo foi analisado a viabilidade financeira, de modo a identificar o perfil de clientes que fornecem prejuízos para a instituição e o perfil de clientes que além de proporcionar o retorno do capital investido também geram lucro para a empresa, garantindo sua saúde financeira e solvabilidade. O modelo de *Credit Scoring* desenvolvido foi eficaz e apresentou resultados satisfatórios para a operadora de saúde, podendo a instituição se beneficiar destas informações para o melhor gerenciamento de riscos.

**Palavras-chave:** *Credit Scoring. Operadora de Saúde. Gestão de Riscos.*

---

<sup>1</sup> DECAT – UFS. e-mail sabrinahours@gmail.com

<sup>2</sup> DECAT – UFS. e-mail marcos@exogenia.com.br

## Python para usuários de R

Lincoln Frias<sup>1</sup>

Patrícia de Siqueira Ramos<sup>2</sup>

Embora a linguagem R seja preferida pelos estatísticos, a linguagem Python é a mais usada em *machine learning* ([www.kaggle.com/surveys/2017](http://www.kaggle.com/surveys/2017)). Por isso, o objetivo deste trabalho é apresentar um panorama da análise de dados em Python, indicando os pacotes essenciais e apontando semelhanças e diferenças em relação ao R. As principais funcionalidades que um estatístico espera estão nos seguintes pacotes: pandas (leitura de dados e operações com dataframes), numpy (operações vetoriais), scipy.stats (distribuições estatísticas), statsmodels (modelos de regressão), pymc3 e pystan (inferência bayesiana) e scikit-learn (*machine learning*). Duas das bibliotecas mais utilizadas para a elaboração de redes neurais são escritas em Python: tensorflow/keras e pytorch. Há diversas opções para a elaboração de gráficos. O pacote matplotlib é capaz de gerar gráficos prontos para publicação, mas bibliotecas como pandas, seaborn e plotnine (uma versão do ggplot2) o utilizam para gerar gráficos esteticamente mais elaborados e de maneira mais simples. No caso de gráficos interativos, há os pacotes de bqplot, altair e plotly. Para a construção de aplicativos para a web, ipywidgets e dash são os equivalentes do shiny. Porém, além da simplicidade da sintaxe, a principal vantagem da Python é o fato de ser uma linguagem de uso geral. Embora ainda não possua certos pacotes estatísticos especializados disponíveis em R, há outros 130 mil pacotes no PyPI (o equivalente do CRAN), em áreas tão diversas quanto edição de áudio e vídeo, construção de sites, astronomia, processamento de texto e manipulações algébricas. Portanto, a Python pode ser uma alternativa ou complemento ao R.

**Palavras-chave:** Python, R, programação.

---

1 Unifal (Universidade Federal de Alfenas, campus Varginha), [lincolnfrias@gmail.com](mailto:lincolnfrias@gmail.com).

2 Unifal (Universidade Federal de Alfenas, campus Varginha), [siqueirapaty@gmail.com](mailto:siqueirapaty@gmail.com).



## Retorno anormal acumulado das ações do setor bancário no período de 1986 a 2017

Matheus Saraiva Alcino <sup>1 5</sup>

Gabriel Rodrigo Gomes Pessanha <sup>2 5</sup>

Lincoln Thadeu Gouvêa de Frias <sup>3 5</sup>

Patrícia de Siqueira Ramos <sup>4 5</sup>

As decisões tomadas nos mercados financeiros pelos grandes investidores desempenham um papel importante na qualidade de vida da população. Tais decisões podem ocasionar oscilações de preço, emprego/desemprego, incentivo a políticas públicas etc. Sendo assim, a partir da ótica de investidores, o presente trabalho busca analisar quais são os fatores que afetam o seu retorno e de que forma o fazem. Para isto, utilizaram-se dados referentes aos cinco maiores bancos brasileiros, de acordo com o seu total do ativo financeiro, ranqueados pelo Banco Central do Brasil. Os dados foram retirados do programa Económica e estão separados trimestralmente, sob o horizonte temporal 1986 – 2017. A análise consiste em aplicar regressão linear múltipla (RLM) para entender as relações do retorno anormal acumulado dessas instituições com os seus determinantes. Os resultados sugerem que parte do comportamento do retorno anormal acumulado de bancos seja explicado pelas variações dos indicadores contábeis e macroeconômicos inclusos no modelo. Assim, as relações obtidas entre a variável de interesse e as variáveis explicativas podem resultar em estratégias eficientes de investimento.

**Palavras-chave:** *Retorno anormal acumulado, mercado acionário, previsão, setor bancário..*

---

<sup>1</sup>ICSA - UNIFAL. e-mail: *osaraivamatheus@gmail.com*

<sup>2</sup>ICSA - UNIFAL. e-mail: *gabrielrgp@yahoo.com.br*

<sup>3</sup>ICSA - UNIFAL. e-mail: *lincolnfrías@gmail.com*

<sup>4</sup>ICSA - UNIFAL. e-mail: *siqueirapaty@gmail.com*

<sup>5</sup>Agradecimento a Fapemig pelo apoio financeiro.

# Retorno Sobre o Patrimônio Líquido em Operadoras de Plano de Saúde - Uma análise através do método Mínimos Quadrados Ordinários

Leonardo Henrique de Andrade B. do Nascimento <sup>1 3</sup>

Ana Carolina de Oliveira <sup>2 3</sup>

Avaliar a rentabilidade das Operadoras de Plano de Saúde (OPS) é extremamente importante, uma vez que a solvência de cada OPS depende do respectivo resultado operacional. Este trabalho tem como objetivo verificar a rentabilidade de uma OPS de grande porte de Minas Gerais. Com isso, pretende-se inferir quais os efeitos que as variáveis macroeconômicas Inflação, Taxa Selic e PIB causam na Rentabilidade sobre o Patrimônio Líquido (ROE). Assim, os dados utilizados compõem uma série temporal anual de 2001 a 2015, retirados dos Anuários: Aspectos econômico-financeiros das OPS, divulgado pela Agência Nacional de Saúde Suplementar (ANS). Com o objetivo de considerar os valores reais das observações utilizou-se o indexador inflacionário IGP-DI. 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 (MQO) através do *software* gratuito Gretl 1.9.90. Os pressupostos acerca dos resíduos da RLM foram testados e respeitados. Em relação ao modelo, a primeira análise mostrou que a variável PIB não era significativa ao nível de 5%. Desta forma, o modelo reduzido contou com as variáveis independentes Inflação e Taxa Selic. O coeficiente de determinação foi de 0,54143, ou seja, das variações ocorridas na ROE, 54,14% delas são explicadas pelas oscilações nas variáveis independentes. Essa estatística foi comprovada pela realização do teste-*F* que se mostrou significativo ao nível de 5%. As variáveis Inflação e Taxa Selic foram significativas ao nível de 5%. O relacionamento entre a Inflação e a ROE é negativo ao passo que Taxa Selic e ROE se relacionam de forma positiva.

**Palavras-chave:** *Operadoras de Plano de Saúde, Regressão Linear Múltipla, ROE.*

---

<sup>1</sup>ICSA - UNIFAL-MG. e-mail: [leonardoandrade.h@outlook.com](mailto:leonardoandrade.h@outlook.com)

<sup>2</sup>ICSA - UNIFAL-MG. e-mail: [ana.coliveira@outlook.com](mailto:ana.coliveira@outlook.com)

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## Uma Análise do IDHM Educação para os Municípios de Minas Gerais mediante Regressão Linear Múltipla

Bárbara Alves Lopes<sup>1 3</sup>

Leonardo Henrique de Andrade Bento do Nascimento<sup>2 3</sup>

Karoline Sanchez Fernandes<sup>2 3</sup>

O Índice de Desenvolvimento Humano Municipal (IDHM) na dimensão Educação é uma composição de variáveis que fornece informação sobre a situação educacional da escolaridade da população adulta e de fluxo escolar da população jovem. Dessa maneira, o objetivo do estudo é avaliar se as variáveis selecionadas possuem alguma relevância na construção do IDHM Educação. A princípio, foram coletadas nove variáveis, provenientes do Censo Demográfico do ano de 2010, fornecidos nos relatórios do Programa das Nações Unidas para o Desenvolvimento – PNUD. Esses dados são referentes aos 853 municípios do estado de Minas Gerais. A análise dos dados foi através Regressão Linear Múltipla (RLM) ajustada pelo método Mínimos Quadrados Ordinários. Toda a análise do modelo linear múltiplo de regressão foi realizada no *software R*. No entanto, o primeiro modelo analisado, obteve estatisticamente apenas cinco variáveis significativas. Diante disso, é testado se há observações influentes, através da Distância de Cook, que identifica o município de Belo Horizonte como um ponto influente, sendo assim retirado da análise. Então, realizou-se um ajuste através do método de *stepwise*, onde foram retiradas duas variáveis independentes. Sendo assim, obtendo o modelo ajustado, com  $R^2$  de 0,8606. Para comprovar essa estatística foi realizado o teste F que rejeitou a hipótese e se mostrou significativo ao nível de 5%. Além disso, foram realizados todos os testes para os resíduos da RLM.

**Palavras-chave:** IDHM Educação, Minas Gerais, Regressão Linear Múltipla.

---

<sup>1</sup> ICSA – UNIFAL. e-mail: barbaraalvesls@yahoo.com.br

<sup>2</sup> ICSA – UNIFAL.

<sup>3</sup> Agradecimento ao CNPq, CAPES, FAPEMIG e UNIFAL pelo apoio financeiro.

## Uma análise dos modelos de apreçamento de ativos aplicados à um banco

Poliana Maria Benelli <sup>1 3</sup>

Gabriel Rodrigo Gomes Pessanha <sup>2 3</sup>

As decisões financeiras geralmente são tomadas em ambientes cercados por incertezas com relação a seus resultados. Quanto maior o risco do ativo, maior será o seu retorno exigido, apesar da postura frente ao risco ser pessoal de cada investidor. Os modelos de precificação de ativos dependem da relação entre retorno e risco para estimar, por exemplo, o capital próprio de uma empresa. O *Capital Asset Pricing Model* (CAPM) é o modelo de precificação mais conhecido, sugerindo que, sob condições de equilíbrio de mercado, a expectativa de retorno sobre determinado ativo acima da taxa livre de risco deve ser proporcional ao seu risco não diversificável, ou risco de mercado, mensurado por  $\beta$ . O modelo APT, desenvolvido por Ross em 1976, se fundamenta na condição de não arbitragem e se coloca como alternativa teórica e empírica ao CAPM. O objetivo deste trabalho é comparar os dois modelos de precificação de ativos, CAPM e APT, para averiguar qual deles é superior na precificação de um ativo do Banco do Brasil. Os dados utilizados são trimestrais e compreendidos entre o quarto trimestre de 1998 ao segundo trimestre de 2017. Os valores coletados e tratados foram importados para o software gratuito Gretl, onde foram realizadas as regressões lineares simples (RLS) e múltiplas (RLM) via método de Mínimos Quadrados Ordinários (MQO). O coeficiente de determinação  $R^2$  do modelo APT mostrou-se superior, porém, utilizando o erro médio, o modelo CAPM apresentou o menor valor e assim, se diz ser superior ao precificar o ativo analisado.

**Palavras-chave:** APT, CAPM, regressão.

---

<sup>1</sup> ICSA - UNIFAL. e-mail polianabenelli@gmail.com

<sup>2</sup> ICSA – UNIFAL.

<sup>3</sup> Agradecimento à FAPEMIG pelo apoio financeiro.

## Uso da metodologia de cópulas na modelagem da dependência de séries temporais de retornos financeiros

Marcela de Marillac Carvalho<sup>1 3</sup>

Thelma Sáfadi<sup>2 3</sup>

As séries temporais de retornos financeiros apresentam características peculiares, que tornam a modelagem da estrutura de dependência desses dados importante para a compreensão de questões em finanças. Neste contexto, a teoria de cópulas constitui uma importante ferramenta na área de econometria e finanças graças a sua flexibilidade para construir funções de distribuição multivariadas que reproduzam dependências não lineares. Conforme esta metodologia, no que se refere à séries temporais, é possível especificar parâmetros de funções cópulas bivariadas de forma fixa, condicionalmente às informações passadas, cuja marginais sigam processos heterocedásticos condicionais da classe ARCH com a inserção de uma parte autorregressiva e de média móvel (ARMA). Utilizando funções cópulas comumente utilizadas na literatura de finanças, Normal, t-Student, Clayton, Gumbel e Frank, especificando marginais com processo com marginais ARMA-APARCH, este trabalho analisou a relação de dependência existente entre o mercado acionário brasileiro e o norte-americano utilizando os respectivos retornos, Ibovespa e S&P500, no período de 04 de janeiro de 2008 a 11 de abril de 2011. Os resultados apontam a existência de uma considerável relação de dependência entre estas economias sugerindo que de forma geral ela é assimétrica, relacionando-se de no período de forma mais acentuada em momentos de alta do que em momentos de baixa do mercado.

**Palavras-chave:** *Cópulas, Finanças, Séries Temporais.*

---

<sup>1</sup>DES - UFLA. e-mail: *marcela-carvalho\_@hotmail.com*

<sup>2</sup>DES - UFLA.

<sup>3</sup>Agradecimento à CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## Uso de polinômios de Bernstein para ajuste de cópulas.

Danielle Gonçalves de Oliveira Prado<sup>1 3</sup>

Lucas Monteiro Chaves<sup>2 3</sup>

Devanil Jaques de Souza<sup>2 3</sup>

Eleanderson Campos Eugênio Filho<sup>2 3</sup>

Chak et al. (2005), Chang et al. (2005, 2007), Curtis e Ghosh (2011), Petrone (1999) e Stadtmuller (1986), entre outros, aplicaram polinômios de Bernstein na realização de regressões com restrição de forma. Wang e Gosh (2012) estudaram a aplicação dos polinômios para o caso de regressão convexa unimodal. Além disso, eles mostraram que o estimador de regressão da base polinomial de Bernstein pode ser obtido como uma solução de um método de mínimos quadrados restritos e, portanto, o estimador pode ser calculado eficientemente usando um algoritmo de programação quadrática. Os polinômios de Bernstein são adequados para a realização de regressões com restrição de forma, em particular para o caso de regressão convexa unimodal. A função de Pickands é justamente convexa e unimodal sendo elemento fundamental na teoria das copulas de valor extremo. Este trabalho está preocupado em explicitar o uso de polinômios de Bernstein na estimação da função de Pickands e estabelecer um novo teste de significância para copulas de valor extremo. Para os resultados necessários sobre teoria das copulas o leitor pode se referir aos textos básicos de Nelsen (2013) e Joe (1997).

**Palavras-chave:** *cópulas, polinômios de Bernstein, função de Pickands.*

---

<sup>1</sup> DAMAT – UTFPR. e-mail: danielle@utfpr.edu.br

<sup>2</sup> DEX – UFLA.

<sup>3</sup> Agradecimento ao CNPq pelo apoio financeiro.

# Engenharia

## **A associação entre os métodos de manutenção e a confiabilidade industrial: uma aplicação da correlação e regressão linear**

**Saymon Ricardo de Oliveira Sousa<sup>1</sup>**

**Cristiane Melchior<sup>2</sup>**

**Ricardo Daher Oliveira<sup>3</sup>**

**Roselaine Ruviaro Zanini<sup>4</sup>**

A prosperidade de uma indústria depende tanto de suas funções centrais quanto as funções técnicas. Com o surgimento da economia globalizada, observou-se um incremento na necessidade dos sistemas em melhor o desempenho produtivo a custos reduzidos, surge então, a confiabilidade, que relaciona-se com a eficiência operacional na ausência de falhas. Ao longo dos anos a manutenção passou a ser uma função estratégica das organizações, encarregada pela disponibilidade operacional dos ativos. A presente pesquisa tem como objetivo estudar o comportamento de um conjunto de variáveis utilizando a análise de correlação e regressão linear, mensurar o grau de associação entre as variáveis e expressar esse efeito matematicamente. Admitiu-se como variável dependente a confiabilidade dos equipamentos e como variáveis independentes os processos de manutenção corretiva e preventiva. O estudo foi realizado em um terminal marítimo portuário localizado no litoral do Maranhão na área embarque de minério de ferro e outros materiais granéis. Foi possível observar que em média para o aumento de uma unidade de variação no processo de manutenção corretiva a confiabilidade do sistema produtivo sofre uma defasagem de aproximadamente 17%. Por outro lado, para o aumento em média no processo de manutenção corretiva a confiabilidade do sistema elava-se em aproximadamente 0,45%. Confirmou-se que a manutenção não planejada impacta diretamente para ineficiência operacional dos equipamentos portuários atenuando a confiabilidade do sistema.

**Palavras-chave:** *Estatística Industrial, Correlação e Regressão, Gestão da Manutenção.*

---

<sup>1</sup> PPGEP - UFSM. e-mail: saymon.ricardo.sousa@gmail.com

<sup>2</sup> PPGEP - UFSM. e-mail: crmelchior@gmail.com

<sup>3</sup> CEUMA. e-mail: ricardo.daher@hotmail.com

<sup>4</sup> PPGEP - UFSM. e-mail: rrzanini63@gmail.com



## Análise de concordância aplicada em modelos para previsão de dispersão de gases inflamáveis

Gislaine Cristina Batistela <sup>1</sup>

Adriana Miralles Schleder <sup>2 4</sup>

Marcelo Ramos Martins <sup>3 4</sup>

Atualmente, é crescente no setor industrial a presença de riscos associados ao vazamento de substâncias inflamáveis cujos efeitos indesejados, como incêndios e explosões, são estimados por modelos que utilizam dinâmica de fluidos computacional. Embora estes modelos sejam promissores, existem poucos estudos de validação dos mesmos. Objetivou-se realizar uma análise de concordância entre os valores da concentração volumétrica do CO<sub>2</sub> mensurados em 10 testes de campo e simulados pelo software FLACS. Ajustou-se modelos de regressão linear simples sem o intercepto ponderando sensores de alturas 0,1 m, 0,6 m e 1,3 m. As estimativas dos coeficientes de regressão foram obtidas pelo método dos mínimos quadrados e complementadas com intervalos de 95% de confiança. Além disso, foi calculado o índice de concordância de Willmott, o qual é uma medida descritiva que reflete a precisão dos valores simulados em relação aos valores mensurados, cujos valores variam de zero, para nenhuma concordância, a 1, para a concordância perfeita. Por meio dos modelos de regressão observou-se que os valores da concentração mensurados e simulados apresentaram concordância em 90% dos cenários na altura de 1,3 m, enquanto que para os sensores de alturas 0,6 m e 0,1 m os valores foram concordantes em 60% dos cenários. Pelo índice de concordância constatou-se que na altura 1,3 m os valores mensurados e simulados têm concordância quase perfeita, ou seja, o menor resultado foi de 0,900. Estes resultados indicaram que a turbulência gerada na região próxima ao solo tem grande influência na estimativa da concentração volumétrica de CO<sub>2</sub>.

**Palavras-chave:** *FLACS, Índice de Willmott, Regressão linear simples.*

---

<sup>1</sup>UNESP - Câmpus de Itapeva. e-mail: [gislaine@itapeva.unesp.br](mailto:gislaine@itapeva.unesp.br)

<sup>2</sup>UNESP - Câmpus de Itapeva.

<sup>3</sup>Departamento de Engenharia Naval e Oceânica - USP

<sup>4</sup>Os autores agradecem o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) para a realização dos ensaios sob o projeto 2013/18218-2.

## Aplicação da combinação de previsões para modelagem de operações em um terminal portuário

Ícaro Romolo Sousa Agostino<sup>1</sup>

Renan Mitsuo Ueda<sup>2</sup>

Maiara de Oliveira Noronha<sup>3</sup>

Adriano Mendonça de Souza<sup>4</sup>

Os modelos de previsões são importantes ferramentas no contexto industrial, auxiliando os processos de tomada de decisão. Uma alternativa para realização de previsões mais precisas é a realização de combinação de vários modelos de previsão. Dessa forma, o objetivo deste estudo foi ajustar um modelo de previsão combinado para prever o comportamento de um sistema de descarga de minério. As observações utilizadas são referentes a um mês de operação, coletadas a cada hora, totalizando uma série temporal 720 observações. Foram verificadas a estacionariedade da série, as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) e suas defasagens, e ajustado modelos concorrentes. Para a combinação das previsões foram adotados os métodos de média aritmética, *Least Squares Weights*, e *Inverse Mean Squared*. A avaliação final dos modelos se deu pela análise das medidas de acurácia MAE, MAPE e o coeficiente de U-Theil. A melhor previsão foi obtida pela combinação dos modelos ARIMA pelo método *Inverse Mean Squared*, apresentando menores valores para as estatísticas de erro, com o MAE de 2,344, MAPE de 2,017. Ao fim do estudo foi possível verificar que os métodos de combinação proporcionaram um aumento da precisão final das previsões, possibilitando o uso do modelo como ferramenta de apoio à decisão.

**Palavras-chave:** *Combinação de previsão, ARIMA, Inverse mean squared.*

---

<sup>1</sup> PPGEF - UFSM. e-mail icaroagostino@gmail.com

<sup>2</sup> PPGEF - UFSM. e-mail renan.mitsuo@hotmail.com

<sup>3</sup> EEMQ - UFSM. e-mail mai.ol.noronha@gmail.com

<sup>4</sup> PPGEF - UFSM. e-mail amsouza.sm@gmail.com

# Confiabilidade de um inspetor eletrônico de garrafas

Paulo César de Resende Andrade <sup>1 3</sup>

Ellen Maria Neves Silva <sup>2 3</sup>

A confiabilidade de um equipamento está associada ao comportamento de falhas em equipamentos. O objetivo deste trabalho foi realizar uma análise da confiabilidade de um inspetor eletrônico de garrafas por meio da modelagem dos tempos até a falha. O procedimento de pesquisa utilizado foi o estudo de caso. Foram disponibilizados quarenta e três dados referentes a falhas ocorridas em dispositivos transportadores pertencentes ao inspetor. Métodos gráficos e analíticos foram utilizados para encontrar a distribuição que melhor modela o conjunto de dados. A análise foi feita utilizando o *software* ProConf. Os testes indicaram a distribuição *Weibull* como a que melhor se ajustou aos dados, com parâmetros de forma  $\gamma$  igual a 1,64 e escala  $\theta$  igual 3,93, determinados pelo método da verossimilhança. O tempo médio até a falha foi de 3,50 minutos. O risco crescente e a confiabilidade decrescente, associados ao fato de ser maior que um, indicam um desgaste do equipamento. Isso corresponde ao início da mortalidade senil, fase final de sua vida útil. A estratégia para esta fase é a manutenção preventiva: a troca antecipa a inevitável quebra.

**Palavras-chave:** *Equipamento, Comportamento de Falhas, Modelagem, Tempo de Vida.*

---

<sup>1</sup>ICT - UFVJM. e-mail: [paulo.andrade@ict.ufvjm.edu.br](mailto:paulo.andrade@ict.ufvjm.edu.br)

<sup>2</sup>Bacharelado em Ciência e Tecnologia - UFVJM.

<sup>3</sup>Agradecimento ao ICT pelo apoio financeiro.

## Delineamento de zonas de manejo via *Topo to raster* em dados assimétricos

Danilo Pereira Barbosa<sup>13</sup>

Lucas Monteiro Bianchi<sup>13</sup>

Eduardo Leonel Bottega<sup>23</sup>

Nerilson Terra Santos<sup>13</sup>

Com o advento da agricultura de precisão, o delineamento de zonas de manejo (ZM) tornou-se prática recorrente quanto à gestão do solo. Tal prática favorece a sustentabilidade agrícola e maximização produtiva. Aliada a esta prática, destaca-se a utilização de informações da condutividade elétrica aparente do solo (CEa), visto que beneficia a produtividade e apresenta baixo custo. Entretanto, suas amostras podem apresentar-se assimétricas, dependendo das condições climáticas. Implicando assim, na ocorrência de *outliers* e necessidade de transformação nos dados. Com isso, objetivou-se a utilização da metodologia *Topo to raster*, visto que independe de tais pressuposições. Para tanto, 160 pontos georreferenciados equidistantes da CEa foram utilizados. Com estes, obteve-se estatísticas descritivas, detectando-se assimetria (2,24) e ocorrência de *outliers*. Posteriormente aplicou-se a metodologia *Topo to raster*. Com o mapa gerado nesta etapa, o algoritmo *fuzzy k-means* viabilizou o delineamento das ZM. Duas ZM foram definidas conforme avaliação dos índices FPI, MPE, Fukuyama-Sugento e xie beni. Estas ZM foram avaliadas quanto à diferença significativa entre si. Para tanto, dentre os modelos lineares mistos avaliados, definiu-se, o modelo de correlação exponencial sem efeito pepita, AIC (1291). As respectivas ZM foram então validadas. Selecionou-se, portanto, 1000 amostras aleatórias destas ZM. E as mesmas, foram verificadas quanto à diferença significativa ( $P < 0,001$ ) entre médias (ZM1:  $\bar{X}_1 = 7$ ; ZM2:  $\bar{X}_2 = 6$ ) da CEa em cada ZM, pelo teste T. Por fim, a metodologia *Topo to raster* facilitou a praticidade no campo (apenas duas ZM) e distinguiu significativamente as ZM entre si, evidenciando sua eficácia no delineamento de ZM.

**Palavras-chave:** *Assimetria, Geoestatística, Topo to raster.*

<sup>1</sup> Departamento de Estatística – UFV. e-mail: danilorv@hotmail.com.

<sup>2</sup> Departamento de Engenharia agrícola – UFSM. e-mail: bottega.elb@gmail.com.

<sup>3</sup> Agradecimento ao CNPq pelo apoio financeiro.

## Facilidades do software ASREML

Alessandra dos Santos <sup>1 2</sup>

Clarice Garcia Borges Demétrio <sup>3</sup>

Renata Alcarde Sermarini <sup>3</sup>

Chris Brien <sup>4</sup>

Com o avanço da tecnologia, a cada dia novas ferramentas são criadas para facilitar o trabalho humano. Na análise de dados é grande o número de softwares criados. A diferença entre eles depende da linguagem de programação, tipo de licença, análises desenvolvidas, entre outras. O ASReml é um poderoso software estatístico em constante desenvolvimento, construído para ajustar modelos lineares mistos considerando-se grandes conjuntos de dados das mais diferentes áreas. Ele possui uma interface com o software R e permite ao usuário o ajuste das mais diversas e complexas estruturas para matrizes de variâncias e covariâncias. O nosso objetivo nesse trabalho é apresentar algumas funções do software ASReml-R bem como a sua abrangência em modelos com informação de parentesco, dependência espacial, efeitos de competição e dados não balanceados.

**Palavras-chave:** *Modelos lineares mistos, Interface R, Ferramenta estatística.*

---

<sup>1</sup>ESALQ/USP. e-mail: [alessandra.s@usp.br](mailto:alessandra.s@usp.br)

<sup>2</sup>Agradecimento ao CNPq pelo apoio financeiro.

<sup>3</sup>ESALQ/USP.

<sup>4</sup>UNISA/Australia.

# Identificação de perfis sensoriais dos consumidores de produtos lácteos via metodologia de múltiplos fatores

Paulo César Ossani <sup>1 4</sup>

Tatiane Carvalho Alvarenga <sup>2 4</sup>

Ariana Campos Frühauf <sup>2 4</sup>

Jéssica Ferreira Rodrigues <sup>3 4</sup>

Renato Ribeiro de Lima <sup>2 4</sup>

Nas diferentes áreas do conhecimento os dados categorizados estão presentes, o que ocorre com frequência nas Ciências dos Alimentos. Dentre as inúmeras técnicas propostas para analisar dados categorizados, a técnica de múltiplos fatores aplicada a tabelas de contingência (MFACT - *Multiple Factor Analysis for Contingency Tables*) se caracteriza por permitir a análise de grupos de variáveis com tamanhos diferentes e de natureza distinta. Um experimento foi realizado no Departamento de Ciências dos Alimentos da Universidade Federal de Lavras, para avaliar o sabor de quatro alimentos: queijo minas, doce de leite, iogurte e mussarela. As categorias de resposta foram avaliadas de acordo com a escala hedônica estruturada de 9 pontos. As amostras avaliadas consistiram em 100 avaliadores de cada sexo nas análises de cada alimento, exceto o queijo minas com 60 avaliadores. Assim este trabalho teve como objetivo verificar se as análises sensoriais feitas nos grupos de alimentos, são similares entre si, ou se são influenciadas pela falta de treinamento em análise sensorial dos provadores. As análises estatísticas foram realizadas utilizando-se o software R. Por meio da metodologia MFACT foi identificada forte similaridade entre os grupos de alimentos, concluindo-se que não houve influência da falta de treinamento sensorial dos provadores, mostrando homogeneidade de modo geral entre eles, além de apontar que independente do produto avaliado sensorialmente os grupos em geral concordaram entre si.

**Palavras-chave:** *MFACT. Pacote MVar.pt . Software R.*

---

<sup>1</sup>DES - UFLA. e-mail: *ossanipc@gmail.com*

<sup>2</sup>DES - UFLA.

<sup>3</sup>DCA - IFMG

<sup>4</sup>Agradecimento a CAPES, CNPQ e FAPEMIG pelo apoio financeiro.

## O alfa de Cronbach deve estar entre 0 e 1?

Carlos Alberto Oliveira de Matos <sup>1 2</sup>

O Alfa de Cronbach é uma medida de confiabilidade com uso relativamente recente em Engenharia. Uma pesquisa no Google Scholar com as palavras “alfa”, “Cronbach” e “Engenharia” produz aproximadamente 52400 resultados, quarenta por cento deles nos últimos cinco anos. Tabelas, devido à sua simplicidade, são muito utilizadas como ferramentas na Engenharia. O uso de tabelas para a interpretação dos coeficientes alfa de Cronbach torna-se mais atraente porque em muitos artigos elas são apresentadas como uma regra de ouro. A maioria das tabelas apresentadas para a classificação da consistência interna assume a suposição errada de que o alfa de Cronbach varia de 0 a 1. Questionários com 3 a 10 itens dicotômicos foram simulados. O fato de que o alfa de Cronbach é baseado nas correlações entre itens explica a inflação aparente de valores negativos que resultaram de valores gerados aleatoriamente e independentemente. No entanto, a distribuição de frequência assimétrica resultante da simulação recomenda que a interpretação da consistência interna por meio de critérios empíricos, tabelados ou não, seja objeto de reflexão crítica e trabalhos adicionais.

**Palavras-chave:** *Alfa de Cronbach, Confiabilidade, Consistência Interna.*

---

<sup>1</sup>Campus Experimental de Itapeva - UNESP. e-mail: [carlos@itapeva.unesp.br](mailto:carlos@itapeva.unesp.br)

<sup>2</sup>Agradecimento à FAPESP pelo apoio financeiro.

## Ensaio clínico e Medicina



## Computação bayesiana aproximada em sistemas de equações diferenciais estocásticas

Maria Cristina Martins <sup>1</sup>

A inferência estatística para dados compartimentais é uma questão recorrente em epidemiologia onde são observados a quantidade de indivíduos infectados ao longo do tempo. Esse tipo de dados é comumente descrito por sistemas de equações diferenciais estocásticas. Tais equações nem sempre possuem solução simples, sendo necessários métodos iterativos para sua resolução e consequentemente para a obtenção de parâmetros que representam as taxas de transmissão ao longo do tempo. Os modelos que descrevem esses processos podem levar a funções de verossimilhança complexas, impossibilitando a utilização de métodos de inferência já conhecidos como os clássicos e os baseados em amostragem. A Computação Bayesiana Aproximada (ABC) é considerada um novo método de inferência com base em resumos estatísticos. Este método de inferência não se baseia na função de verossimilhança e depende exclusivamente de simulações numéricas do modelo. Algumas de suas vantagens são a eficiência computacional e inferência estatística bayesiana uma vez que, funciona bem para pequenas amostras e possibilita incorporar informações a priori em um parâmetro. A Computação Bayesiana Aproximada foi utilizada em um modelo de equações diferenciais estocásticas para realizar a predição do número de indivíduos infectados por HIV em Cuba.

**Palavras-chave:** *Dados compartimentais, Livre de verossimilhança, Resumo estatístico.*

---

<sup>1</sup>FZEA - USP. e-mail: *m.martins@usp.br*

## Ferramentas gráficas para auxílio em seleção de variáveis

Lucas Ragiotto <sup>1 3</sup>

Jacqueline Domingues <sup>2 3</sup>

Luzia Aparecida Trinca <sup>4</sup>

Em problemas de regressão linear, na busca por um modelo parcimonioso, o pesquisador pode se deparar com adversidades, por exemplo, a existência de colinearidade entre as regressoras, dificultando a seleção de variáveis. Dessa forma, com a implementação do pacote `mplot` no *software* R, que utiliza o método bootstrap e uma inspiração do método *fence* (Jiang, J. et al., 2009, Stat. Probab. Lett., 79(5), 625-629), pode-se, graficamente e interativamente, analisar com detalhes a estabilidade de modelos e a inclusão de co-variáveis, permitindo também o estudo de suas interações dado os efeitos principais. Esta ferramenta gráfica utiliza o critério de informação generalizado como procedimento para seleção de modelos, uma medida formada pelo componente que reflete a “falta de ajuste” e pelo componente que reflete a complexidade do modelo ponderado por uma constante de penalidade. Os critérios de informação de *Akaike* e Bayesiano, utilizam essa constante como sendo 2 e  $\log(n)$ , respectivamente. O `mplot` permite a realização de uma inspeção num intervalo real da constante, auxiliando o pesquisador na análise de incorporação de variáveis explicativas. Fez-se uma aplicação deste procedimento para modelar o peso de bebês nascidos prematuramente em Botucatu entre 1995-2000, utilizando-se uma amostra de dados que contém treze covariáveis. O estudo da estabilidade de vários modelos alternativos e a investigação do gráfico de inclusão de variáveis levou a um modelo final com três regressoras, sendo elas idade gestacional em semanas, gênero e realização ou não de pré-natal. Explorou-se também os efeitos de interação duas a duas dessas covariáveis, mas nenhum deles foi selecionado.

**Palavras-chave:** *Bootstrap, Critério de informação generalizado, Regressão linear.*

---

<sup>1</sup>PPG Biometria, IBB, UNESP. e-mail: [lucasragiotto@gmail.com](mailto:lucasragiotto@gmail.com)

<sup>2</sup>PPG Biometria, IBB, UNESP. e-mail: [jacquelinedomingues.unesp@gmail.com](mailto:jacquelinedomingues.unesp@gmail.com)

<sup>3</sup>Agradecimento à CAPES pelo apoio financeiro.

<sup>4</sup>Depto. de Bioestatística, IBB, UNESP.

# Full Bayesian Significance Test para Cópula PVF

Murilo Cantoni <sup>1</sup>

Jose S. Romeo <sup>2 3</sup>

Adriano Polpo <sup>4</sup>

A família arquimediana de dois parâmetros da cópula Power Variance Function (PVF) inclui as cópulas de Clayton, Gumbel e Gaussiana Inversa como casos especiais ou limites, oferecendo assim uma abordagem unificada para ajustar essas cópulas amplamente utilizadas. A ideia deste trabalho é implementar o Full Bayesian Significance Test (FBST) para testar hipóteses precisas e, portanto, uma avaliação rigorosa dos casos particulares das cópulas de Clayton, Gumbel e Gaussiana Inversa no modelo de cópula PVF geral. Propomos o uso do FBST para comparar modelos encaixados pois acreditamos que esta é uma alternativa interessante uma vez, diferentemente de outras técnicas de comparação de modelos, é de fato um teste de hipóteses. Nós ilustramos a utilidade da metodologia usando dados de tempos até apendicectomia para gêmeos adultos no Australian NH&MRC Twin Registry.

**Palavras-chave:** *Cópulas Arquimedianas, Dependência, Full Bayesian Significance Test, Análise de Sobrevivência Multivariada.*

---

<sup>1</sup>DEs-UFSCar and ICMC-USP. e-mail: [murilo.ufscar@gmail.com](mailto:murilo.ufscar@gmail.com)

<sup>2</sup>Department of Mathematics, University of Santiago, Chile.

<sup>3</sup>Institute of Natural & Mathematical Sciences, Massey University, New Zealand.

<sup>4</sup>DEs-UFSCar.

## Joint model para dados longitudinais e multi-estado: Uma aplicação para câncer de próstata

Felipe Barletta <sup>1</sup>

Dr. Paulo Canas <sup>2</sup>

Dra. Isolde Previdelli <sup>3</sup>

Joint models para dados longitudinais e de sobrevivência, são uma ferramenta poderosa que levam em conta estes dois tipos de dados simultaneamente em um único modelo, permitindo inferir sobre a dependência e associação entre um biomarcador longitudinal e eventos no tempo. Esses modelos são úteis para estudos no campo da saúde que visam a compreensão da doença, considerando o seu desenvolvimento ao longo do tempo e a quantidade de tempo até o paciente atingir um estado absorvente. Os mais utilizados, que resultam de uma combinação de um modelo longitudinal e análise de sobrevivência, não permitem monitorar a ligação entre o biomarcador longitudinal e as transições entre os múltiplos estados da doença até o estado absorvente. Para entender melhor esta ligação entre o biomarcador longitudinal e as transições entre os múltiplos estados, usou-se um modelo conjunto que combina o longitudinal e o Markov multi-estado. Uma aplicação é apresentada em dados de câncer de próstata. Os parâmetros do modelo são estimados pelo método da máxima verossimilhança em dois estágios: (i) os efeitos fixos e aleatórios são estimados; e (ii) essas estimativas são usadas, para vincular o modelo longitudinal com o modelo multi-estado de Markov, permitindo a mensuração do impacto para o risco de morte, considerando covariáveis, em cada transição entre os estados da doença ao longo do tempo. Desta forma, o modelo é capaz de avaliar a trajetória do biomarcador, definir os riscos das transições entre estados de saúde e quantificar o impacto da dinâmica do PSA em cada intensidade de transição.

**Palavras-chave:** *Joint model, Modelo longitudinal, Modelo multi-estado de Markov, Câncer de próstata..*

---

<sup>1</sup>PBE - UEM. e-mail: [felipe.e.barletta@gmail.com](mailto:felipe.e.barletta@gmail.com)

<sup>2</sup>DEST - UFBA. e-mail: [paulocanas@gmail.com](mailto:paulocanas@gmail.com)

<sup>3</sup>DES - UEM. e-mail: [isoldeprevidelli@gmail.com](mailto:isoldeprevidelli@gmail.com)

## Mais atividade física e menos estresse: estudo sobre o estilo de vida de universitários

Carlos Henrique dos Santos Silva<sup>1</sup>

Paola Michelli Ferreira Rosa<sup>1</sup>

Cléber Luz<sup>1</sup>

Marcos Santos Oliveira<sup>1</sup>

Juliana Kátia Silva<sup>2</sup>

Este trabalho visa analisar a influência da dor, dos sintomas de ansiedade, depressão e estresse e nível de atividade física no estilo de vida de universitários. Participaram 44 universitários saudáveis que não estavam em tratamento farmacológico, psicológico ou fisioterapêutico durante a coleta. O desfecho primário foi o estilo de vida (EV) captado através do questionário Estilo de Vida Fantástico. As variáveis preditoras foram nível de atividade física (BEACHE modificado), sintomas de ansiedade, depressão e estresse ( $EADS_{21}$ ) e intensidade da dor pela escala visual analógica-EVA. Foi realizada uma regressão linear múltipla considerando  $\alpha=0,05$ . O modelo de regressão que contemplou o nível de atividade física, estresse e depressão foi estatisticamente significativo [ $F(3,40) = 7,624$ ,  $p=0,000$ ]. As variáveis preditoras foram responsáveis por 36,4% da variação do EV. Contudo, os sintomas de depressão não foram capazes explicar significativamente a variação do estilo de vida ( $t=13,0608$ ,  $p=0,501$ ), apenas o nível de atividade física ( $t=2,180$ ,  $p=0,035$ ) e o estresse ( $t=2,634$ ,  $p=0,011$ ). Com os resultados, indica-se que os universitários aumentem o nível de atividade física e diminuam a sobrecarga de estresse para melhorar o EV dos universitários.

**Palavras-chave:** *Estilo de vida, Universitários, Regressão Linear Múltipla.*

---

<sup>1</sup>Mestrando em Biometria e Estatística Aplicada - UFRPE. e-mail: [chsestatistica@gmail.com](mailto:chsestatistica@gmail.com)

<sup>1</sup>ICS - UFBA. e-mail: [paola\\_michelli@hotmail.com](mailto:paola_michelli@hotmail.com)

<sup>1</sup>ICS - UFBA. e-mail: [cleberluz@outlook.com](mailto:cleberluz@outlook.com)

<sup>1</sup>Docente do Departamento de Estatística e Ciências Atuariais - UFS. e-mail: [marcos@exogenia.com.br](mailto:marcos@exogenia.com.br)

<sup>2</sup>Docente do Departamento de Estatística e Ciências Atuariais - UFS. e-mail: [juliana.ks@ufs.br](mailto:juliana.ks@ufs.br)

# Modelo de regressão quase-beta multivariado

Ricardo Rasmussen Petterle <sup>1</sup>

Wagner Hugo Bonat <sup>2</sup>

Cassius Tadeu Scarpin <sup>3</sup>

Modelos de regressão multivariados são atrativos, pois permitem a análise conjunta de múltiplas respostas além de outras vantagens na análise dos dados. Nesse contexto, o presente artigo propõe um novo modelo de regressão para análise de variáveis respostas limitadas multivariada, chamado por modelo de regressão quase-beta multivariado. O modelo é especificado usando apenas suposições de primeiro e segundo momentos e o método de estimação proposto combina as funções de estimação quase-escore e Pearson para estimação dos parâmetros de regressão e dispersão, respectivamente. A principal vantagem da abordagem proposta é não precisar assumir uma distribuição de probabilidade para as variáveis respostas, além da modelagem conjunta da matriz de correlação entre as respostas. Outra vantagem desta especificação é a simplicidade e rapidez do algoritmo de estimação, que pode ser resumido em um simples algoritmo do tipo Newton-score. Foram delineados dois estudos de simulação. O primeiro foi feito para avaliar o comportamento do algoritmo NORTA na simulação de variáveis aleatórias beta correlacionadas. Na sequência, delineou-se o segundo estudo de simulação para explorar o comportamento dos estimadores sob diferentes cenários. Os resultados mostraram que em todos os cenários os estimadores são consistentes, não viciados e eficientes. O modelo foi motivado pela aplicação em um conjunto de dados referente ao percentual de gordura corporal, que foi dividido em cinco regiões do corpo e representam as variáveis respostas. Os resultados mostraram que o modelo multivariado apresentou um melhor desempenho na análise dos dados, quando comparado às análises feitas separadamente para cada variável resposta (modelo univariado).

**Palavras-chave:** *Variáveis respostas limitadas multivariada, Intervalo unitário, Dados correlacionados, Estudo de simulação, Algoritmo NORTA.*

---

<sup>1</sup>Universidade Federal do Paraná - UFPR, Setor de Ciências da Saúde, Curso de Medicina, Curitiba, PR, Brasil. e-mail: [ricardopetterle@ufpr.br](mailto:ricardopetterle@ufpr.br)

<sup>2</sup>Universidade Federal do Paraná - UFPR, Departamento de Estatística, Laboratório de Estatística e Geoinformação, Curitiba, PR, Brasil. e-mail: [wbonat@ufpr.br](mailto:wbonat@ufpr.br)

<sup>3</sup>Universidade Federal do Paraná - UFPR, Departamento de Administração Geral e Aplicada, Curitiba, PR, Brasil. e-mail: [cassiusts@gmail.com](mailto:cassiusts@gmail.com)

## Modelo Logístico para curvas de crescimento com estrutura heterocedástica.

Magali Teresopolis R. Amaral <sup>1</sup>

Carlos R. Padovani <sup>2</sup>

Antônio C. Cicogna <sup>3</sup>

Em diversas áreas de conhecimento os modelos não lineares são ajustados levando-se em consideração que os erros são normais homocedásticos. No entanto, o pressuposto de igualdade entre as variâncias nem sempre é satisfeito e a presença de heterocedasticidade pode ser constatado em diversas situações práticas (Louzada *et al.*, 2010). Neste sentido, o trabalho objetiva utilizar o método de estimação de máxima verossimilhança no modelo logístico com erros normais heterocedásticos para ajustar curvas de crescimento da tensão máxima desenvolvida (TD) do músculo papilar isolado em ratos em função dos diferentes níveis de doses sequenciais de cálcio  $d = (0.5, 1.0, 1.5, 2.0, 2.5, 3.0)$  mM administradas após a cirurgia para indução de Estenose Aórtica supravalar (EAo).

O modelo logístico é definido por  $y_{id} = \alpha(1 + e^{-\gamma d})^{-\delta}$ , em que  $y_{id}$  é a média da  $i$ -ésima unidade experimental avaliada na  $d$ -ésima dose,  $\alpha$  representa o crescimento assintótico do modelo,  $\gamma$  é a taxa que determina a eficiência do crescimento da TD e  $\delta$  é o parâmetro que determina o ponto que ele passa a crescer com menor eficiência.

Para explicar o fenômeno de crescimento, considerou-se a estrutura de covariância proposta por Harvey (1976),  $\Sigma = \text{diag}(\sigma^2 d^\lambda)$ , que está associada aos vetores dos erros heterocedásticos com estrutura multiplicativa.

Assim, após avaliar a qualidade do modelo, pode-se inferir que este foi adequado, pois forneceu estimativas consistentes para o entendimento da TD em função das doses administradas.

**Palavras-chave:** *Curvas de crescimento, heterocedástica multiplicativa, modelo de regressão não linear, Verossimilhança.*

---

<sup>1</sup>DEXA-UEFS. e-mail: [mteresopolis@hotmail.com](mailto:mteresopolis@hotmail.com)

<sup>2</sup>IBB-UNESP.

<sup>3</sup>FMB-UNESP.

## Modelo não linear misto no estudo do ganho de peso de ratos

Jair Wyzykowski <sup>1</sup>

Luiz Paulo Campos Patrício <sup>2</sup>

O objetivo do presente trabalho foi avaliar, através de um modelo não linear misto, o efeito do uso de café (*Coffea arabica*), no controle de pesos de ratos submetidos a dietas alimentares com e sem extrato aquoso do mesmo. Os dados de peso foram obtidos em um experimento realizado no Departamento de Farmácia da Universidade Federal de Minas Gerais, para avaliar seis dietas alimentares. O delineamento experimental foi o inteiramente casualizado, com dez repetições. Os pesos dos ratos foram medidos semanalmente, durante quinze semanas, após quarta semana do nascimento. Através da análise gráfica exploratória do comportamento dos dados inerentes aos tratamentos, decidiu-se pelo uso do modelo não linear misto logístico, ajustando-se as curvas para cada tratamento com o uso do *software* R e a biblioteca *nlme*. Foram testadas três estruturas de variância-covariância e a necessidade de efeito aleatório nos parâmetros com o uso dos critérios de informação de Akaike e Bayesiano, além do teste da razão de verossimilhança. O modelo misto não linear logístico foi adequado para descrever o ganho de peso em ratos quando adotada a matriz positiva definida diagonal para a estrutura de variância-covariância dos efeitos aleatórios e, ao parâmetro assíntota, é acrescentado o efeito aleatório. Ademais, todas as pressuposições inerentes ao modelo foram satisfeitas.

**Palavras-chave:** *efeito aleatório; modelo misto; regressão não linear.*

---

<sup>1</sup>Professor Adjunto do Centro de Ciências Agrárias, Ambientais e Biológicas da Universidade Federal do Recôncavo da Bahia. Área de Estatística Experimental Aplicada às Ciências Agrárias. e-mail: [jair.stat@gmail.com](mailto:jair.stat@gmail.com)

<sup>2</sup>Estudante do Curso de Agronomia/Bolsista Voluntário. e-mail: [lpaulocp@hotmail.com](mailto:lpaulocp@hotmail.com)



## Modelos de regressão de Cox e fragilidades Gama, Inversa Gaussiana e t: Um estudo comparativo aplicado a dados de retinopatia

Rodolfo Crystian Pereira SILVA <sup>1</sup>

Alisson de Lima BRITO <sup>1</sup>

Ana Patrícia Bastos PEIXOTO <sup>1</sup>

Ednário Mendonça da SILVA <sup>1</sup>

Ricardo Alves de OLINDA <sup>1</sup>

Tiago Almeida de OLIVEIRA <sup>2</sup>

A Diabetes Mellitus caracteriza-se pela elevação da glicose no sangue e segundo a OMS, 422 milhões de adultos ao redor do mundo têm esta doença, quando não controlada pode provocar lesões progressivas na retina e nos vasos sanguíneos dos olhos, que pode levar a cegueira. Uma dessas lesões é a Retinopatia Diabética. No estudo de análise de sobrevivência geralmente não se conhece todas as variáveis que podem afetar o tempo até a ocorrência do desfecho. Muitas vezes, quando sabe-se que existem muitos outros fatores que não foram considerados no estudo e a população não pode ser considerada homogênea, o modelo de regressão de Cox com fragilidade surge como uma alternativa ao modelo de riscos proporcionais de Cox. O primeiro é um modelo de efeito aleatório, onde o efeito aleatório (fragilidade) tem um efeito multiplicativo na função de risco de base. Este efeito aleatório explica a dependência nos modelos de fragilidade. Neste trabalho, objetivou-se estudar o tempo até a cegueira total de pacientes com Retinopatia Diabética que faziam ou não um tratamento à *laser*, para isso utilizou-se o modelo de regressão de Cox e os modelos de fragilidade de Cox com as distribuições Gama, Gaussiana e t, para modelar a fragilidade. Após a comparação entre os modelos no intuito de encontrar o que melhor descrevia os dados do tempo até a cegueira por retinopatia, selecionou-se o modelo com fragilidade Inversa Gaussiana (menor AIC 1533,81 e concordância de 0,84). As variáveis selecionadas foram Tratamento e Risco (escore de risco para um olho).

**Palavras-chave:** *Heterogeneidade, Sobrevivência, Riscos proporcionais.*

---

<sup>1</sup>DE - UEPB

<sup>2</sup>DE - UEPB. e-mail: [tiagoestatistico@gmail.com](mailto:tiagoestatistico@gmail.com)

## O uso de modelos de fração de cura bivariados na análise dos tempos de sobrevida e remissão de pacientes diagnosticados com câncer de pele

Ricardo Puziol de Oliveira <sup>1</sup>

Jorge Alberto Achcar <sup>1</sup>

A modelagem e análise dos dados médicos com sobreviventes de longo prazo em que o objetivo principal é o tempo até que um evento de interesse ocorra, é de grande importância em estudos médicos. Nestes estudos, é possível a presença de dois ou mais tempos de vida associados à mesma unidade, como os tempos de níveis de deterioração ou os tempos de reação a um tratamento em pares de órgãos, como pulmões, rins, olhos ou ouvidos. Também é possível a presença de uma fração de cura onde é necessário modelar o tempo de vida com sobreviventes de longo prazo. Neste trabalho, é introduzido, sob uma abordagem bayesiana, um modelo de fração de cura de mistura bivariado com base na distribuição geométrica bivariada de Basu-Dhar para investigar os sobreviventes de longo prazo diagnosticados com câncer de pele do tipo melanoma, uma vez que este é o tipo de câncer mais frequente no Brasil correspondendo a 30% de todas as neoplasias malignas registradas no país e tem alta possibilidade de metástase.

**Palavras-Chave:** *análise de sobrevivência, estudos médicos, inferência Bayesiana, melanoma, métodos MCMC.*

---

<sup>1</sup>Departamento de Medicina Social. Faculdade de Medicina de Ribeirão Preto - FMRP/USP. e-mail: [rpuziol.oliveira@gmail.com](mailto:rpuziol.oliveira@gmail.com); [achcar@fmrp.usp.br](mailto:achcar@fmrp.usp.br)

## Teste da herdabilidade multivariada aplicado a dados de famílias

Alex de Oliveira Ribeiro <sup>1</sup>

Daniel Furtado Ferreira <sup>2</sup>

Júlia Maria Pavan Soler <sup>3</sup>

Doenças complexas como a síndrome metabólica, cardiopatias e Alzheimer, constituem um importante problema de saúde pública. De forma geral, essas doenças estão ligadas a distúrbios nos mecanismos de controle de múltiplas variáveis que se expressam de maneira conjunta. A partir do modelo misto poligênico, Blangero et al. (2013) chegaram à primeira expressão analítica conhecida para o teste da herdabilidade de uma única variável. Porém, metodologias capazes de testar simultaneamente a herdabilidade presente em um conjunto de variáveis, como é o caso das doenças complexas, ainda não foram desenvolvidas. Diante do exposto, o presente trabalho foi realizado com o objetivo principal de desenvolver um teste estatístico para a herdabilidade multivariada. Em virtude das relações de dependência entre as variáveis, presentes nesse contexto, utilizou-se a técnica de componentes principais de herdabilidade - PCH para a geração de novas variáveis independentes. A estatística de teste obtida apresentou forma analítica relativamente simples. Sua avaliação foi conduzida por meio de simulação Monte Carlo com 5000 simulações para cada cenário configurado em função das herdabilidades paramétricas, correlações e estrutura familiar. Além disso, a metodologia também foi aplicada a dados reais relacionados à síndrome metabólica pertencentes ao Projeto Corações de Baependi. Os resultados obtidos demonstraram que sob a hipótese nula (ausência de herdabilidade multivariada) o teste proposto foi capaz de controlar, de forma eficiente, as taxas de erro tipo I e, sob a hipótese alternativa (presença de herdabilidade multivariada) seu poder foi elevado.

**Palavras-chave:** *Componentes Principais de Herdabilidade, Doenças Complexas, Teste da Razão de Verossimilhanças.*

---

<sup>1</sup>DES - UFLA. e-mail: [alex.ribeiro@des.ufla.br](mailto:alex.ribeiro@des.ufla.br)

<sup>2</sup>DES - UFLA

<sup>3</sup>IME - USP

# Uma extensão de modelos bivariados usando modelos de fração de cura em estudos médicos

Ricardo Puziol de Oliveira <sup>1</sup>

Jorge Alberto Achcar <sup>1</sup>

Modelos de fração de cura são amplamente utilizados em estudos médicos quando existem pacientes não-suscetíveis ao evento de interesse. Uma vez que há presença desses pacientes, independente da natureza do tempo observado, em geral, assume-se uma distribuição contínua sem incorporar a fração de cura ou o modelo de riscos proporcionais de Cox para realizar a modelagem. Entretanto, em muitas situações a medida de tempo é discreta e a presença da fração de cura induz uma limitação aos modelos clássicos de análise de sobrevivência, especialmente no caso bivariado. No presente trabalho, ilustramos uma nova abordagem usando modelos bivariados discretos com fração de cura aplicados a dados médicos com uma fração de cura bem delimitada. Os resultados mostraram que os modelos de fração de cura devem ser considerados para uma melhor análise por oferecerem uma precisão melhor. Além disso, os resultados inferenciais sob o enfoque Bayesiano mostraram uma boa performance com baixo custo computacional mostrando que o uso de modelos discretos pode ser uma boa alternativa para a análise de dados de sobrevivência com fração de cura.

**Palavras-Chave:** *análise de sobrevivência, estudos médicos, inferência Bayesiana, diabetes, métodos MCMC.*

---

<sup>1</sup>Departamento de Medicina Social. Faculdade de Medicina de Ribeirão Preto - FMRP/USP. e-mail: [rpuziol.oliveira@gmail.com](mailto:rpuziol.oliveira@gmail.com); [achcar@fmrp.usp.br](mailto:achcar@fmrp.usp.br)

## Ensino

# Análise multi-estado da evasão de estudantes da USP

Fabiana Arca Cruz Tortorelli <sup>1</sup>

Danilo Kroll Mantello <sup>2</sup>

Juliana Cobre <sup>3</sup>

O curso de Bacharelado em Matemática Aplicada e Computação Científica do ICMC da USP do campus de São Carlos tem enfrentado grande número de evasão discente. Dessa forma, a análise de dados torna-se uma ferramenta necessária para avaliar quais fatores estão relacionados com o abandono do curso. Para isso, foi realizada análise descritiva dos dados que contêm informações socioeconômicas e acadêmicas de 375 estudantes que ingressaram no curso entre os anos 2000 e 2016, além de abordagem tradicional de análise de sobrevivência e modelos multiestados ou riscos competitivos para estimar a evasão discente e os fatores relacionados. Verificou-se que na estatística descritiva e exploratória, algumas variáveis foram significativas para os estudos. Além disso, o método de Kaplan-Meier (abordagem clássica) forneceu funções de sobrevivência superestimadas. Porém, com a função de incidência acumulada, uma estimação alternativa que calcula a probabilidade de ocorrência de um evento levando em conta a presença de riscos competitivos foi observado que probabilidade de evasão está sempre acima da conclusão no decorrer do tempo, além de que mais de 50% dos alunos ingressantes desde 2000 não concluíram o curso. A partir dos resultados é possível observar que o problema da evasão no curso em questão é relevante para entender e buscar novos estudos com outras variáveis ou métodos que expliquem ainda mais a evasão discente.

**Palavras-chave:** *Evasão discente, Modelos multi-estados, Riscos competitivos.*

---

<sup>1</sup>PIPGEs - UFSCar/USP. e-mail: [fabii@usp.br](mailto:fabii@usp.br)

<sup>2</sup>USP. e-mail: [mantello@usp.br](mailto:mantello@usp.br)

<sup>3</sup>ICMC/USP. email: [jucobre@icmc.usp.br](mailto:jucobre@icmc.usp.br)

## **Características socioeconômicas e desempenho no ENEM em Ciências Exatas**

**Joyce Dalline Silva Andrade<sup>1</sup>**

**Lorena França Andrade<sup>1</sup>**

**Elizeu Junio Dantas Alves<sup>1</sup>**

**João Marcos Melo Santos<sup>1</sup>**

**José Rodrigo Santos Silva<sup>1,2</sup>**

O trabalho proposto visa estudar o desempenho do estudante do estado de Sergipe no ENEM, na área Ciências Exatas e Tecnológicas, relacionando o desempenho dos estudantes com suas condições socioeconômicas. Desta forma, pretende-se entender quais os fatores que contribuem para um melhor desempenho dos estudantes. Para a análise foram utilizados os microdados do ENEM 2016, disponibilizados pelo INEP. Foram utilizadas as técnicas de Análise de Componentes Principais (com Rotação Varimax) e Modelos Lineares Generalizados (Família Binomial com ligação logit) na análise dos dados. Admitiu-se como desempenho satisfatório os estudantes que possuíram nota acima da mediana (450 pontos). Foram selecionadas 6 componentes, com a manutenção de 63,48% da variação total. Os resultados apontaram que a componente principal PC1 (residentes em Aracaju, pai ou mãe com ensino superior, alta renda familiar e que nunca estudaram em escola pública), PC3 (candidatos brancos ou amarelos) e PC5 (estudantes oriundos do ensino regular, que estudaram em turno diurno e que nunca abandonaram ou reprovaram) funcionaram como fatores protetores ao mau desempenho na prova (Odds Ratio, respectivamente igual a 0,67, 0,95 e 0,87), enquanto que as PC2 (candidatos com maior idade, casados ou em união estável, que já concluíram o ensino médio e que trabalham), PC4 (candidatos cujos pais possuem ensino fundamental ou médio apenas) e PC6 (pessoas do sexo feminino e pessoas sem acesso a computador ou internet) foram verificadas como fatores de risco ao desempenho satisfatório em Ciências Exatas e Tecnológicas (Odds Ratio, respectivamente, 1,05, 1,03 e 1,47).

**Palavras-chave:** *Desempenho, Microdados, ENEM.*

---

<sup>1</sup> Departamento de Estatística e Ciência Atuarial – Universidade Federal de Sergipe

<sup>2</sup> Autor correspondente, e-mail: rodrigo.ufs@gmail.com

## **Classificação dos cursos da Universidade Federal de Sergipe a partir da taxa de desistência no Sisú**

**Elizeu Júnio Dantas Alves<sup>1</sup>**

**João Marcos Melo Santos<sup>1</sup>**

**Joyce Dalline Silva Andrade<sup>1</sup>**

**Lorena França Andrade<sup>1</sup>**

**José Rodrigo Santos Silva<sup>1,2</sup>**

O Sistema de Seleção Unificado (Sisú) funciona como uma espécie de leilão, onde o candidato deve escolher, por ordem de preferência, até duas opções entre as vagas ofertadas pelas instituições participantes. É permitido ao candidato, durante o período de inscrição, modificar suas opções quantas vezes julgar conveniente, devendo este ficar atento à nota de corte do curso de seu interesse para garantir a permanência na lista. Após o término do período de inscrições, o SISU/MEC realiza a chamada regular dos aprovados, onde as vagas não aproveitadas na chamada regular são divulgadas em uma lista de espera. Se novas vagas surgem, por desistência dos candidatos classificados, os suplentes são transformados em classificados. Na Universidade Federal de Sergipe observou-se uma grande quantidade de desistências, que resultaram em 9 chamadas de excedentes em 2017, com a convocação de 1842 candidatos no total. O presente trabalho buscou identificar em quais cursos houveram maior e menor taxa de desistência em 2017, onde, através de técnicas de Análise de Agrupamento, buscou-se classificar os cursos quanto a taxa de desistência. Em 36% dos cursos ofertados nos Campus Aracaju e São Cristóvão o número de desistentes foi maior que o número de vagas ofertadas. Observou-se que os maiores percentuais de desistência estão diretamente associados com o conhecimento dos candidatos acerca da profissão, sendo os cursos de Teatro (86%), Ciência da Religião (86%) e Ciências Atuariais (82%) os cursos que apresentaram maior taxa de desistência, e os cursos com menor desistência foram Astronomia (6%), Medicina (9%) e Direito (12%).

**Palavras-chave:** *Sisú, Taxa de Desistência, Excedentes.*

---

<sup>1</sup> Departamento de Estatística e Ciência Atuarial – Universidade Federal de Sergipe

<sup>2</sup> Autor correspondente, e-mail: rodrigo.ufs@gmail.com



## Construção de Escalas para Avaliação de Proficiência com Base na Teoria da Resposta ao Item

Tânia Robaskiewicz Coneglian Fujii <sup>1 4</sup>

Aparecida Donizete Pires de Souza <sup>2</sup>

Adriano Ferreti Borgatto <sup>3</sup>

A Teoria da Resposta ao Item (TRI) consiste em um conjunto de modelos matemáticos, que representam a probabilidade de um indivíduo dar uma certa resposta a um item, como função dos parâmetros dos item e da sua habilidade (ou traço latente). Neste trabalho, utilizou-se o modelo logístico unidimensional de três parâmetros (ML3), com o objetivo de estimar a proficiência em conteúdos matemáticos básicos, de ingressantes nos cursos da área de exatas. Este é um dos modelos mais utilizados para itens dicotômicos (corrigidos como certo ou errado), caracterizado pelos parâmetros de discriminação, de dificuldade e de acerto ao acaso. Devido à complexidade do modelo, para o procedimento de inferência sobre os parâmetros e habilidades, sob o paradigma clássico ou bayesiano, se faz necessário o uso de métodos numéricos. Neste caso, utilizou-se o amostrador de Gibbs, algoritmo da classe dos Métodos de Monte Carlo via Cadeia de Markov (MCMC), implementado via software OpenBUGS (Bayesian Inference Using Gibbs Sampling), direcionado para análise bayesiana de modelos complexos. Como resultados são apresentados formas para construção de escalas, baseadas na Teoria da Resposta ao Item, com ênfase na área educacional, envolvendo a forma de definição dos pontos na escala, como são estabelecidas as distâncias entre eles e como cada ponto deve ser interpretado. Além disso, a construção e interpretação de uma escala para avaliação de proficiência em conteúdos matemáticos básicos.

**Palavras-chave:** *Avaliação da Proficiência em Conteúdos Matemáticos Básicos, Construção de Escalas, Inferência Bayesiana, Teoria de Resposta ao Item.*

---

<sup>1</sup>PosMAC - UNESP - Presidente Prudente. e-mail: [taniaconeglian@gmail.com](mailto:taniaconeglian@gmail.com)

<sup>4</sup>Agradecimento à CAPES pelo apoio financeiro.

<sup>2</sup>UNESP - Presidente Prudente.

<sup>3</sup>UFSC.

## Ensino de Medidas descritivas por meio de Resolução de Problemas

Cleomar Viater Júnior<sup>1 3</sup>

Luiz Felipe Zitkoski<sup>2 3</sup>

Matheus Henrique Dal Molin Ribeiro<sup>3 3</sup>

Os Parâmetros Curriculares Nacionais (PCNs) e as Diretrizes Curriculares da Educação Básica do Paraná (DCE), especificamente para o terceiro e quarto ciclos do ensino fundamental, procuram direcionar os alunos ao raciocínio estatístico e probabilístico, por meio de circunstâncias em que o aluno consiga desenvolver eloquência quanto a informações que envolvam o raciocínio probabilístico e combinatório. Para que isso seja possível, tais documentos propõem abordagens metodológicas para o ensino de estatística, dentre as quais pode-se destacar a Resolução de Problemas. Ela fornece ao aluno liberdade para desenvolver o raciocínio, desprendendo-se da matemática rotineira e desinteressante que valoriza a reprodução. O aluno sozinho toma consciência de que necessita entender o problema, montar uma estratégia de solução e verificar se ela funciona. Então, expõe-se aqui um estudo onde aplicou-se a metodologia Resolução de Problemas como proposta de ferramenta para o ensino de estatística, em especial de medidas de tendência central, a saber, média, moda e mediana. Pensando que o professor serve de intermédio entre conhecimentos e alunos, desenvolvem-se atividades para o levante da aplicação da metodologia em sala de aula. A atividade foi realizada no Colégio Estadual Professor Agostinho Pereira, na cidade de Pato Branco, em turmas do oitavo ano do ensino fundamental, com um total de 80 alunos aproximadamente, por bolsistas do Programa Institucional de Bolsas de Iniciação à Docência (Pibid). Deste modo, a partir do desenvolvimento da atividade, observou-se que tal abordagem mostrou-se eficiente no auxílio do ensino dos conceitos estatísticos trabalhados, incitando a participação e discussão dos problemas propostos como forma de fixar os conteúdos abordados.

**Palavras-chave:** *Ensino de estatística, Medidas de Tendência Central, Resolução de Problemas.*

---

<sup>1</sup>Universidade Tecnológica Federal do Paraná - Pato Branco. E-mail: [cleomarviater7@gmail.com](mailto:cleomarviater7@gmail.com)

<sup>2</sup>Universidade Tecnológica Federal do Paraná - Pato Branco. E-mail: [uizfelipezitkoski@gmail.com](mailto:uizfelipezitkoski@gmail.com)

<sup>3</sup>Universidade Tecnológica Federal do Paraná - Pato Branco. E-mail: [mrribeiro@utfpr.edu.br](mailto:mrribeiro@utfpr.edu.br)

<sup>3</sup>Esta pesquisa contou com apoio financeiro da CAPES através de concessão de bolsas de Iniciação à docência pelo Programa Institucional de Bolsas de Iniciação à Docência – PIBID.

# Ferramenta interativa para o estudo de modelos básicos de probabilidade

Andryas Waurzenczak <sup>1</sup>

## Resumo

Este trabalho é uma ferramenta interativa com o objetivo de simplificar a visualização de modelos teóricos de probabilidade, visto em cursos introdutórios de estatística básica. Foi desenvolvido utilizando o software R, com auxílio do pacote shiny. Atráves da ferramenta é possível mostrar em tempo real o comportamento das distribuições ao se alterar os parâmetros correspondentes. Além disso, pode-se calcular a probabilidade entre um intervalo especificado pelo usuário, verificando à área abaixo da curva nos casos contínuos e a massa de probabilidade nos casos discretos. Por último, tem-se a descrição da função de distribuição/densidade de probabilidade e os dois primeiros momentos.

**Palavras-chave:** *Ferramenta interativa, ensino, modelos de probabilidade.*

---

<sup>1</sup>UFPR. e-mail: *Andryaas@gmail.com*

## Influência das monitorias na aprovação dos discentes em Matemática 1 e 2

Leonardo Biazoli <sup>1</sup>

Alice Duarte <sup>2</sup>

Cláudia Adam Ramos <sup>2</sup>

A alta reprovação nas disciplinas de Matemática 1 e 2 não é exclusiva do Bacharelado Interdisciplinar em Ciência e Economia (BICE) da Universidade Federal de Alfenas - UNIFAL-MG - *campus* Varginha, conforme Lopes (1999), as disciplinas de matemática são aquelas que mais reprovam, sejam no ensino superior ou não. Nesse sentido, o programa de monitoria busca auxiliar na diminuição desses índices de forma efetiva e dinâmica. Diante disso, o presente trabalho buscou investigar a influência da monitoria no desempenho dos discentes, da UNIFAL-MG *campus* Varginha, nas disciplinas de Matemática 1 e 2, nos dois semestres do ano de 2017. Os dados foram fornecidos pela CRCA/UNIFAL-MG (Coordenadoria de Registros e Controle Acadêmico) e analisados computacionalmente por meio do *software* R. Para analisar o primeiro semestre de 2017 foi calculado a probabilidade condicional de um discente ser aprovado em Matemática 1 e ter frequentado trinta monitorias, e também, a probabilidade de ser aprovado e não comparecer nessas monitorias. Obteve-se que a primeira probabilidade foi de aproximadamente 0,7273, enquanto a segunda foi de aproximadamente 0,0926. Analogamente, analisou-se o segundo semestre de 2017, encontrando uma taxa de 0,5455 para aqueles que frequentaram as monitorias e foram aprovados na disciplina, e para a outra probabilidade, de ser aprovado na Matemática 2 e não ter frequentado todas as monitorias, encontrou uma taxa de 0,25. Dessa forma, pode-se observar que os discentes frequentes nas monitorias obtiveram probabilidade maiores de serem aprovados quando comparados aos não frequentes.

**Palavras-chave:** *Monitoria, Probabilidade, Desempenho.*

---

<sup>1</sup>Universidade Federal de Alfenas - UNIFAL-MG. e-mail: [leonardobiazoli19@gmail.com](mailto:leonardobiazoli19@gmail.com)

<sup>2</sup>Universidade Federal de Alfenas - UNIFAL-MG.

## LECA – Pacote Estatístico Educacional em R com Interface Gráfica

Ana Carolina Abreu Barbosa<sup>13</sup>

Deyse Márcia Pacheco Gebert<sup>23</sup>

O pacote educacional LECA, desenvolvido em R, vem sendo projetado desde 2016 dentro do programa de iniciação científica na Universidade Estadual de Ponta Grossa, visando integrar métodos Estatísticos em uma interface gráfica para manipulação de dados e fins didáticos. Sua versão atual disponibiliza os seguintes métodos para variáveis aleatórias qualitativas e quantitativas: tabelas de distribuição de frequências (incluindo frequências acumuladas), medidas descritivas (posição, dispersão, assimetria e curtose) e gráficos (histogramas, polígonos, ogivas, barra e pizza). Todas as medidas e gráficos são gerados a partir da tabela de distribuição de frequências e essas funções podem ser executadas através da interface amigável do pacote, condição que pretende a inclusão da ferramenta R em salas de aulas de cursos dentro e fora da área da Estatística, independente da familiarização dos alunos com a linguagem. Funcionalidades adicionais a partir dos métodos anteriormente expostos, como manipulação de arquivos e cálculos de probabilidades, já estão em fase de desenvolvimento e futuramente espera-se que o pacote acople um vasto ramo de opções para seus usuários.

**Palavras-chave:** *Estatística Computacional, Estatística Descritiva, Software R.*

1 DEINFO – UEPG. e-mail anacbse@gmail.com

2 DEMAT – UEPG. e-mail dmpgebert@uepg.br

3 Agradecimento à UEPG pelo apoio financeiro.

## Notas de matemática no vestibular $\times$ desempenho na disciplina de pré-cálculo no ensino superior: existe mesmo uma relação?

Luiz R. Nakamura <sup>1</sup>

André W. Zibetti <sup>1</sup>

Andreia Zanella <sup>1</sup>

Yohana T. Hoffmann <sup>2</sup>

Neste trabalho tentamos descrever a relação entre as notas de matemática no vestibular (MTM) e as notas na disciplina de pré-cálculo (PC) de alunos da Universidade Federal de Santa Catarina. As notas de MTM referem-se às notas/pontuação obtida no tema matemática do concurso vestibular, aplicada em nível nacional, com grau de dificuldade equivalente a cada concurso de vestibular. Notas de PC se referem às notas dos alunos obtidas em uma avaliação (nos primeiro e segundo semestres de 2017) que envolve resolver questões de matemática. Notamos que os modelos de regressão usuais (modelos lineares clássicos, modelos lineares generalizados e modelos aditivos generalizados) não são suficientes para descrever a variável notas de PC a partir das notas de MTM. Assim, os modelos aditivos generalizados para locação, escala e forma (GAMLSS) (baseado na distribuição  $t$  de Student assimétrica), foram utilizados para tal fim. Como esperado, a partir do modelo ajustado, observamos que quanto maior a nota de MTM obtida pelo aluno, maior a média de nota de PC. Ainda, verificamos que a variabilidade nas notas de PC é muito menor nos grupos de alunos que estão nas extremidades (isto é, alunos com notas ou muito altas ou muito baixas na disciplina) em relação as notas que os alunos obtiveram em MTM, do que aqueles que obtiveram notas medianas (de quatro à seis pontos) no vestibular. Finalmente, a maior assimetria presente em notas de PC está na faixa onde os alunos obtiveram nota de MTM entre dois e quatro pontos.

**Palavras-chave:** *Ensino, Funções de suavização, GAMLSS, t de Student assimétrica.*

---

<sup>1</sup>INE – UFSC. e-mail: [luiz.nakamura@ufsc.br](mailto:luiz.nakamura@ufsc.br)

<sup>2</sup>Programa de Pós-Graduação em Educação Científica e Tecnológica – UFSC

## Programação orientada a objetos no R: relações ecológicas para a coleção de *cards* ilustrados em classes S3 e S4

Hektor Dannyel Vieira Brasil

Walmes Marques Zeviani

Um dos paradigmas mais utilizados em programação é a orientação por objetos (POO). A linguagem de programação R possui três abordagens para desenvolvimento de POO: S3, S4 e R6. A S3 foi a primeira metodologia de classes implementada, sendo mais simples e abrangente. A S4 é mais rígida e limitada, mantendo relações hierárquicas entre as classes. A metodologia R6 é a mais recente, desenvolvida para ser mais leve e rápida. O objetivo do projeto é desenvolver, de forma simples e dirigida ao ensino, a lógica de POO com as metodologias de classe S3 e S4 na linguagem R e produzir um pacote R que sirva de exemplo. O contexto considerado foi o preenchimento de álbuns de *cards* por uma comunidade simulada de colecionadores. Dois tipos de relação ecológica foram considerados para a troca entre colecionadores, para avaliar qual é mais favorável para a conclusão do álbum: interesse bilateral ou unilateral. Observou-se que a implementação em S3 é mais simples mas requer mais passos para simular todas as relações. A implementação em S4 é mais laboriosa porém goza de um desenvolvimento mais robusto. No fim, a comunidade de colecionadores com cooperação unilateral teve um tempo menor para a conclusão do álbum.

**Palavras-chave:** *Simulação Montecarlo, Classes e Métodos.*

## ShinyExams: uma Interface Shiny para o pacote exams

Gian de Lima Santos <sup>1</sup>

Fernando de Pol Mayer <sup>2</sup>

O pacote **exams** para R auxilia na criação e correção de exames, baseando-se na produção de exercícios com conteúdo dinâmico. As questões podem ser criadas em  $\text{\LaTeX}$  ou Markdown e incluir códigos R para construir as questões de forma aleatória. A ideia do projeto ShinyExams é permitir que qualquer pessoa crie exames à partir deste pacote, porém sem obrigatoriamente ter conhecimento em programação. Considerando esta necessidade, nasceu o projeto de criar uma interface que facilita a criação de exames empregando funções do pacote **exams**. Para isso, foram utilizados pacotes que auxiliam na criação da interface do usuário, como **Shiny** e **ShinyDashboard**. A interface ShinyExams permite que o usuário consulte um banco de questões (próprio ou compartilhado), visualize cada uma delas e decida quais incluir em uma prova. Depois disso, podem ser geradas  $n$  provas aleatórias, com folha de rosto onde as respostas serão marcadas, e gabarito (opcional). A interface também permite que o usuário crie novas questões, que já ficam armazenadas no banco de questões. Também foi criado um sistema de *login* para que usuários decidam se querem compartilhar seus exercícios. Com a interface, os professores tem uma maior rapidez na criação de exames. Os educadores não precisam criar todo o código base, caso fosse feito em linguagem R. Também podem visualizar e utilizar exercícios criados por outros docentes, promovendo um ambiente na qual os usuários podem não apenas gerar exercícios, mas também compartilhar entre eles.

**Palavras-chave:** *exams*, *Interface*, *Shiny*.

---

<sup>1</sup>LEG/DEST/UFPR. e-mail: [gian.lima@live.com](mailto:gian.lima@live.com)

<sup>2</sup>LEG/DEST/UFPR.



## Uma década de Ações Afirmativas na UFSCar

Maria Sílvia de Assis Moura <sup>1</sup>

Em 2006 o Conselho Universitário da UFSCar aprovou um programa de ações afirmativas para ingresso nos cursos de graduação. O programa foi montado com a perspectiva de reservar um percentual para os estudantes que concluíram inteiramente todo o ensino médio em escola pública e 35% dessa parcela para estudantes negros, pretos ou pardos. O programa previa uma implementação gradual, durante os anos de 2007, 2008 e 2009, a parcela destinada às ações afirmativas foi de 20%, depois 30% para 2010, 40% em 2011 e 50% a partir de 2012. Em 2012 foi promulgada uma lei que instituiu o programa de ações afirmativas em todo sistema federal de ensino superior, a diferença entre o programa nacional e o programa da UFSCar foi a reserva por renda. Antes de 2013 as modalidades de ingresso eram Ampla Concorrência, Escola Pública e Escola Pública Negros, com a lei passamos a ter, Ampla Concorrência, Escola Pública com renda superior a 1,5 s.m. per capita, Escola Pública com renda igual ou inferior a 1,5 s.m. per capita, Escola Pública Pretos, Pardos ou Indígenas com renda superior a 1,5 s.m., e Escola Pública Pretos, Pardos ou Indígenas com renda igual ou inferior a 1,5 s.m. Esse trabalho apresenta uma análise estatística da situação e do desempenho dos estudantes da UFSCar ao longo desses 10 anos. Utilizamos análise descritiva uni e bi-variada e não constatamos diferenças de desempenho ou de situação entre as modalidades de ingresso.

**Palavras-chave:** *Ações Afirmativas, Comparação de Médias, Desempenho Acadêmico.*

---

<sup>1</sup>DEs - UFSCar. e-mail: [msilvia@ufscar.br](mailto:msilvia@ufscar.br)

# Uma interface para cálculo de produtos atuariais

Walef Machado de Mendonça <sup>1</sup>

Yagho Lemes Carvalho <sup>2</sup>

Ana Carolina Orrico <sup>2</sup>

Pedro Henrique Rezende Oliveira <sup>2</sup>

Danilo Machado Pires <sup>2</sup>

Nas últimas décadas os produtos atuariais, como o seguro de vida e as anuidades, tem passado por enormes mudanças. O grande aumento no poder computacional, obtido nos últimos anos, tem possibilitado o desenvolvimento de novos e inovadores produtos. Como intuito de fornecer uma base educacional, o principal objetivo deste trabalho é integrar teoria dos produtos do ramo vida, demonstrando a grande variedade de produtos atuarias, com uma abordagem mais prática e interativa através de uma interface. A interface e todas as análises foram desenvolvidas utilizando o programa gratuito e de código aberto R. Tal linguagem foi utilizada por meio da interface RStudio, que além de tornar o ambiente de programação mais intuitivo, possibilita o uso da biblioteca Shiny. Atualmente a interface possibilita o cálculo dos principais produtos do ramo vida, sendo eles: Seguro de vida (temporário e vitalício), anuidades (temporária e vitalícia) e seguro dotal e dotal misto. Com a possibilidade do usuário escolher os parâmetros utilizados para o cálculo do produto. A interface oferece de maneira rápida e intuitiva a possibilidade de replicar os cálculos aliando o poder computacional da linguagem de programação R à interatividade na alteração dos parâmetros e variáveis a serem utilizadas, se apresentando assim, como uma importante ferramenta para o estudo da matemática atuarial do ramo vida.

**Palavras-chave:** *Matemática atuarial, anuidades, interface educacional, seguro de vida*

---

<sup>1</sup>ICSA/UNIFAL-MG e-mail walefm2@gmail.com

<sup>2</sup>ICSA/UNIFAL-MG

# Um aplicativo shiny para modelos lineares generalizados

Cayan Atreio Portela Bárcena Saavedra <sup>1</sup>

Cristian Marcelo Villegas Lobos <sup>2</sup>

Recentes avanços tecnológicos, automatizações e esta era de grandes informações tem sido responsáveis por diversas mudanças em análises e visualizações de dados. Uma dessas mudanças, caracteriza-se no uso de plataformas interativas e dinâmicas para tais análises. O recente crescimento no uso de gráficos dinâmicos e interativos baseados em web pages, demonstra haver um interesse significativo na criação dessas plataformas. Um dos motivos para tal crescimento, é a maneira de realizar visualizações e análises de dados, que não se limita mais a um ambiente estático, possibilitando um maior leque na maneira de apresentar e investigar os dados.

O software R (R Core Team, 2017) permite ao usuário realizar análises estatísticas com um forte apoio computacional. Com seu avanço diversos pacotes e plataformas foram sendo desenvolvidos proporcionando ao usuário um maior leque de ferramentas. A plataforma shiny (Chang et al. 2015) é um pacote desenvolvido que basicamente fornece ao usuário a possibilidade de criar aplicações web de forma fácil e intuitiva, combinando a capacidade computacional do R com visualizações interativas, sem ser necessário o conhecimento de outras linguagens de programação tais como CSS, Javascript e HTML. A criação de uma aplicação shiny não é complexa, porém exige um conhecimento intermediário de programação na linguagem R. Podem ser criados documentos em que o usuário pode interagir e visualizar em tempo real as alterações em gráficos e tabelas. Com essa ferramenta, podem ser criados aplicativos úteis em diversas áreas e para diversos fins.

O presente trabalho tem como objetivo demonstrar um aplicativo shiny para modelos lineares generalizados, destacando sua utilidade para usuários-fim sem afinidade em programação e no ensino de conceitos relacionados à estatística, bem como, a vantagem de interações simultâneas, visando um melhor entendimento multi-dimensional da análise destes modelos.

**Palavras-chave:** *Modelos lineares generalizados, shiny, R.*

---

<sup>1</sup>LCE - ESALQ. e-mail: [cayan@usp.br](mailto:cayan@usp.br)

<sup>2</sup>LCE - ESALQ. e-mail: [clobos@usp.br](mailto:clobos@usp.br)

## Epidemiologia e Saúde Pública

## Ajuste de Modelos de Regressão na Avaliação da Função Pulmonar de Alunos das Escolas Públicas do Município de Campina Grande-PB

Alisson de Lima Brito <sup>1 3</sup>

Ricardo Alves de Olinda <sup>1 3</sup>

Danillo Barros Cordeiro <sup>1 3</sup>

Danielle Franklin de Carvalho <sup>2 3</sup>

Tiago Almeida de Oliveira <sup>1 3</sup>

Em muitas situações práticas nem sempre os pressupostos do modelo de regressão clássica são alcançados. Sendo assim, precisa-se de uma abordagem mais flexível que alcance dados de natureza contínua com comportamento positivo assimétrico. Com efeito, os Modelos Lineares Generalizados (MLG's), por ser versátil, permite que a variável resposta se adeque a esse comportamento, sendo as distribuições normal inversa e gama adequadas para essa modelagem. Portanto, tem-se como objetivo ajustar modelos onde a variável independente é de natureza contínua com comportamento positivo assimétrico utilizando os MLG's, mostrando aos pesquisadores das ciências da saúde a importância da utilização dessa metodologia. Os dados utilizados são referentes à função pulmonar de adolescentes no município de Campina Grande - PB. As variáveis associadas a este estudo se ajustaram bem aos modelos assimétricos propostos, constatando que as distribuições normal inversa e gama podem ser utilizadas para ajustar modelos com comportamento assimétrico na variável resposta.

**Palavras-chave:** *Dados assimétricos, Modelos de regressão, Saúde pública.*

---

<sup>1</sup>Departamento de Estatística - UEPB. e-mail: *ricardo.estat@yahoo.com.br*

<sup>2</sup>Departamento de Enfermagem - UEPB.

<sup>3</sup>Agradecimento à Universidade Estadual da Paraíba pelo apoio financeiro.

## Ajuste de modelos de regressão para o número de casos notificados por dengue no estado da Paraíba

Alisson de Lima BRITO <sup>1</sup>

Andreza Jardelino da SILVA <sup>2</sup>

Ana Patrícia Bastos PEIXOTO <sup>1</sup>

Tiago Almeida de OLIVEIRA <sup>3</sup>

A dengue é atualmente um dos principais problemas enfrentados pelo sistema de saúde e pela população em si, uma vez que pode levar um indivíduo à morte e é de difícil controle. Infelizmente a única forma de combate à essa doença é o controle de seu mosquito transmissor, o *Aedes Aegypti*. A técnica de Modelos Lineares Generalizados (MLG) pode ser de grande valia para podermos entender a forma de propagação da dengue entre os municípios da Paraíba. O objetivo deste trabalho foi verificar possíveis fatores socioeconômicos e termopluiométricos que possam afetar o número de casos de dengue notificados em municípios do estado da Paraíba conjuntamente com a escolha da melhor distribuição que se ajusta aos dados por meio de um ajuste de um MLG. Os dados foram no site DATASUS <sup>4</sup>, IBGE <sup>5</sup>, AESA<sup>6</sup> e CLIMATE<sup>7</sup>. Para a modelagem desse conjunto de dados, utilizou-se as distribuições de Poisson, Quasi-Poisson, Binomial Negativa. Comparou-se os modelos quanto a distribuição utilizada e procedeu-se com o gráfico seminormal de probabilidade com envelope simulado para os três modelos para verificação da qualidade de ajuste. O modelo Quasi-Poisson foi selecionado e a interpretação do modelo final escolhido foi que, para cada aumento de uma unidade de milímetros cúbicos de chuva acumulada tem-se um aumento de 0,07% na quantidade de casos notificados. É possível ver também que a medida que se aumenta a renda decai o caso de dengue em 7% e para cada grau celsius de temperatura aumenta-se em 30% o número de notificações.

**Palavras-chave:** *Quasi-Poisson, Socioeconômico, MLG.*

---

<sup>1</sup>DE - UEPB.

<sup>2</sup>Esalq - USP

<sup>3</sup>DE - UEPB. e-mail: [tiagoestatistico@gmail.com](mailto:tiagoestatistico@gmail.com)

<sup>4</sup><http://datasus.saude.gov.br/>

<sup>5</sup><http://www.ibge.gov.br/home/>

<sup>6</sup><http://www.aesa.pb.gov.br/>

<sup>7</sup>[www.climate-data.org](http://www.climate-data.org)

## Análise de dados de área aplicada a variável malária em Moçambique

Cláudio Francisco Chipenete<sup>1 3</sup>

Marcelo Silva de Oliveira<sup>1 3</sup>

Gisela H. Nunes Chipenete<sup>2 3</sup>

A estatística espacial na análise de dados de área tem sido uma ferramenta bastante útil em diversas áreas de pesquisa. O interesse centra-se em descobrir respostas para as seguintes perguntas: qual é a estrutura ou distribuição do fenómeno na região em estudo? Existiria algum padrão definido ou tendência entre as diversas áreas nessa região? O objetivo principal neste trabalho foi implementar duas técnicas estatísticas, descritas na literatura para estimar o índice de Moran global e local, comparando os resultados. Foram utilizados dados do inquérito demográfico de saúde de Moçambique de 2011, de casos de malária em 146 distritos disponibilizados pela *The DHS Program*. Os resultados mostraram que não há diferenças significativas para as estimativas dos índices global e local de Moran pelo método tradicional e pelo método desenvolvido por Chen (2013). Foi possível identificar padrões na região sul e extremidade norte do país com taxas baixas, sugerindo a influência de algum fator não descrito no estudo. Localmente, distritos de Maputo, Matola e Beira tiveram maiores taxas de incidência. Maior concentração e mobilidade populacional nesses distritos provavelmente explica esse resultado.

**Palavras-chave:** estatística espacial, dados de área, malária

---

1 DES – UFLA. e-mail [claudiochipenete@hotmail.com](mailto:claudiochipenete@hotmail.com)  
2 DES – UFLA  
3 PAG – UFLA

## Instruções para submissão de trabalhos

Walmes Marques Zeviani<sup>1 3</sup>

Fernando de Pol Mayer<sup>2 3</sup>

Wagner Hugo Bonat<sup>2 3</sup>

### 1 Introdução

Este texto é um modelo para a construção de “Trabalho Completo” a ser submetido na 63<sup>a</sup> Reunião Anual da RBras que ocorrerá no DEST/UFPR, Curitiba - PR, no período de 23 à 25 de maio de 2018.

### 2 Instruções para o trabalho

- 1) O resumo simples e o abstract devem ser feitos em arquivos separados do resumo expandido. O resumo simples deve conter, no máximo, 250 palavras, enquanto que o abstract tem limite de 50 palavras. Ambos devem ser submetidos em dois formatos: .tex ou .docx, e .pdf.
- 2) O resumo expandido deve conter as seguintes seções: 1 – Introdução, 2 – Material e métodos, 3 – Resultados e discussões, 4 – Conclusões e 5 – Bibliografia. Utilizar CAIXA ALTA apenas na primeira letra de cada título;
- 3) Figuras e tabelas devem ser centralizadas e numeradas, sendo que o título da tabela deve ser colocado acima da mesma, ao passo que para figura, o título deve estar especificado abaixo;
- 4) A numeração das fórmulas deverá estar à direita;
- 5) As referências deverão ser listadas em ordem alfabética no final do trabalho. No caso de artigos, destacar o período em questão e no caso de livros publicados, destacar o título do livro (Seguir os exemplos apresentados no final desse documento ).

### 3 Apresentação oral

Para apresentação oral, os autores deverão submeter um trabalho completo na forma de resumo expandido contendo todas as seções anteriormente citadas. Dependendo do número

---

1 DEST – UFPR. e-mail walmes@ufpr.br

2 DEST – UFPR.

3Agradecimento ao CNPq pelo apoio financeiro.



de trabalhos completos inscritos, alguns deles poderão ser convidados para apresentação como pôsteres.

#### 4 Apresentação em pôster

Para apresentação em pôster, os autores deverão submeter um resumo simples de uma página.

#### 5 Referências

- [1] ASSOCIAÇÃO BRASILEIRA DE HEREFORD E BRAFORD. **Hereford – Carne de qualidade tipo exportação**. Disponível em: <http://www.hereford.com.br/>. Acesso em: 13 de fevereiro de 2011.
- [2] BARBIN, D. **Planejamento e análise estatística de experimentos agronômicos**. Arapongas: Editora Midas Ltda. 2003. 194 p.
- [3] HINDE, J.; DEMÉTRIO, C. G. B. Overdispersion: models and estimation. **Computational Statistics & Data Analysis**. Elsevier. v. 27, n. 2, p. 151-170, 1998.
- [4] PESCIM, R. R. **A distribuição beta generalizada semi-normal**. 2009. 124 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica), Universidade de São Paulo, Piracicaba, 2009.

## **Análise de sobrevivência espacial de indivíduos diagnosticados com AIDS no município de Campinas em três coortes: 1980-1990, 1996-2000 e 2001-2005.**

**Liciane Vaz de Arruda Silveira<sup>1</sup>**

**Thiago Santos Mota<sup>1</sup>**

**Ana Carolina Nunes Mafra<sup>2</sup>**

**Maria Rita Donalísio<sup>3</sup>**

Este trabalho tem como objetivo avaliar a sobrevida e o risco de óbito por qualquer causa de adultos com AIDS no município de Campinas-SP, em função de variáveis sócio-demográficas, comportamentais, clínicas e localizações espaciais. Os dados foram obtidos do Sistema de Informação de Agravos de Notificação (SINAN) e do Sistema de Informação sobre Mortalidade (SIM) de Campinas no período de 1980 a 2014. As análises de sobrevida foram realizadas em três coortes que se referem aos anos de diagnóstico/notificação (1980-1990, 1996-2000 e 2001-2005). Modelos de riscos proporcionais com efeitos espaciais aditivos foram ajustados aos dados possibilitando obter os riscos espaciais de óbito por AIDS no município. Na coorte 1, a variável idade maior que 30 anos foi associada a maior mortalidade; na coorte 2, idade maior que 30 anos, drogas injetáveis, tuberculose e apresentar menor contagem de células CD4 foram associadas a maior mortalidade. Na coorte 3, o modelo foi estratificado pela variável imunossupressão grave, devido esta variável ferir a pressuposição de riscos proporcionais, e o modelo indicou que sexo masculino e baixa escolaridade estão associados a maior mortalidade. A variação espacial para o risco de morte foi significativa para a coorte 1 e não significativa para as coortes 2 e 3. O estudo mostra contextos diferentes nas diversas etapas da epidemia em Campinas, indicando que mesmo em uma cidade com o programa de IST-AIDS bem consolidado é necessário acompanhamento contínuo para conhecer os fatores de risco de morte por AIDS, devido às mudanças da epidemia ao longo do tempo.

**Palavras-chave:** *modelos aditivos generalizados, modelo de Cox, coordenadas geográficas.*

<sup>1</sup> Departamento de Bioestatística, IBB/UNESP, SP, Brasil.

<sup>2</sup> Instituto Israelita de Ensino e Pesquisa, Hospital Israelita Albert Einstein. SP, Brasil.

<sup>3</sup> Departamento de Saúde Coletiva, FCM/UNICAMP. SP, Brasil.

## Análise do Deslocamento de Pacientes Submetidos à Hemodiálise em Curitiba em 2016

Maria Helena Santos de Oliveira<sup>1</sup>

Guilherme Fernando Hathy<sup>1</sup>

Walmes Marques Zeviani<sup>2</sup>

Wagner Hugo Bonat<sup>2</sup>

O tratamento para nefropatias muitas vezes é um processo contínuo. O deslocamento de pacientes gera custos tanto monetários quanto de qualidade de vida. Assim, estudar a distribuição das percursos dos pacientes para os hospitais abre possibilidades de investimentos e melhorias.

O estudo teve como objetivo investigar o deslocamento de pacientes em tratamento de nefropatias que necessitam de hemodiálise, e ainda propor a otimização do sistema de distribuição dos pacientes entre os hospitais de Curitiba. A pesquisa foi desenvolvida utilizando dados públicos disponibilizadas pelo SIASUS (Sistema de Informações Ambulatoriais do SUS).

A análise apontou que apenas 34% dos pacientes estavam sendo atendidos no hospital mais próximo da sua residência, ou seja, 66% percorriam uma distância maior do que a necessária considerando uma alocação baseada no percurso mais curto.

**Palavras-chave:** *Bioestatística, Estatística Espacial, Nefrologia.*

---

<sup>1</sup>Estudantes - Bacharelado em Estatística UFPR

<sup>2</sup>DEST - UFPR

## Análise espacial dos casos de dengue do município de Campina Grande - PB no ano de 2016

Pablo Lourenço Ribeiro de Almeida <sup>1 6</sup>

Jorge Garcia Filho <sup>2 6</sup>

Elias Silva de Medeiros <sup>3 6</sup>

Kelly Pereira de Lima <sup>4 6</sup>

Renato Ribeiro de Lima <sup>5 6</sup>

No Brasil, a dengue é considerada um sério problema de saúde pública, devido principalmente as epidemias constantes e contínuas a cada ano, sendo possível encontrar a presença do mosquito em todo território nacional. Segundo o Ministério da Saúde, no ano de 2016, até a semana epidemiológica 52 (3/1/2016 a 31/12/2016), foram registrados 1.500.535 casos confirmados de dengue no país. Desse total, foram registrados 360 casos confirmados no município de Campina Grande - PB. Em estudos epidemiológicos, como por exemplo os casos de dengue, a identificação e quantificação de padrões na ocorrência de doenças fornecem os primeiros passos para uma maior compreensão e, possivelmente, o controle das mesmas. Ao considerar a localização de cada caso da doença é possível identificar um comportamento espacial no local onde esse evento acontece. Dessa forma, os métodos da estatística espacial oferecem um meio de considerar a informação da localização desses eventos, com a finalidade de detectar e quantificar padrões espaciais da doença em alguma região específica (Ex: países, estados, municípios, etc.). Portanto, o objetivo desse trabalho foi utilizar a estatística espacial, mais especificamente a análise de dados de área, para compreender os casos de dengue no município de Campina Grande - PB no ano de 2016. Ao utilizar o índice  $I$  de Moran, obteve-se uma autocorrelação espacial significativa. Além disso, identificou-se de forma positiva os bairros com maiores riscos de contaminação da doença.

**Palavras-chave:** *Autocorrelação Espacial, Índice  $I$  de Moran, Epidemiologia.*

---

<sup>1</sup>DES - UFLA. e-mail: [pablo.lourenco@posgrad.ufla.br](mailto:pablo.lourenco@posgrad.ufla.br)

<sup>2</sup>DES - UFLA. e-mail: [jorgegarcia@posgrad.ufla.br](mailto:jorgegarcia@posgrad.ufla.br)

<sup>3</sup>DES - UFLA. e-mail: [eliasmedeiros@posgrad.ufla.br](mailto:eliasmedeiros@posgrad.ufla.br)

<sup>4</sup>DES - UFLA. e-mail: [kelly\\_plima@posgrad.ufla.br](mailto:kelly_plima@posgrad.ufla.br)

<sup>5</sup>DES - UFLA. e-mail: [rrlima@des.ufla.br](mailto:rrlima@des.ufla.br)

<sup>6</sup>Agradecimento a CAPES, CNPq e FAPEMIG pelo apoio financeiro.

# Análise espaço-temporal de casos novos anuais de hanseníase no Estado do Paraná (1980-2013), por meio de equações diferenciais estocásticas parciais

Vanessa Ferreira Sehaber <sup>1 2</sup>

Paulo Justiniano Ribeiro Jr <sup>3</sup>

Elias Teixeira Krainski <sup>4</sup>

## 1 Introdução

Em estudos epidemiológicos, é desejável compreender e explicar o processo saúde-doença nos indivíduos e em populações que vivem em uma dada região e um período de interesse. Como premissa geral, tais estudos assumem que os padrões de morbi-mortalidade e saúde não ocorrem de forma aleatória em populações humanas, mas sim em padrões ordenados que refletem causas subjacentes [BRASIL 2006].

A ocorrência de casos novos de uma doença trás preocupação aos órgãos da saúde pública, pois há indicativos de que a população está sob risco. Por essa razão, os casos novos precisam ser detectados e controlados ainda em seus estágios iniciais [PENNA 2011].

Com o intuito de desenvolver estratégias para monitoramento e tomada de ações, o Ministério da Saúde estabeleceu a Lista Nacional de Notificação Compulsória - Portaria n.º 204 de 2016, na qual estão elencadas as doenças que precisam ser comunicadas às autoridades de saúde pública por meio do SINAN - Sistema Nacional de Agravos Notificáveis [BRASIL 2009]. Em tal lista, se encontra a hanseníase, uma doença milenar e que ainda causa estigmas sociais nos tempos atuais.

A Organização Mundial de Saúde (OMS) tem particular interesse na erradicação da hanseníase, dado o seu potencial de incapacitação. Segundo o último levantamento da OMS, o Brasil é o segundo país com maior registros de casos novos de hanseníase no mundo (232.857), ficando atrás da Índia (135.485) [OMS 2017].

Segundo dados parciais do SINAN/2017, há apenas três estados com detecção baixa de hanseníase (Rio Grande do Sul, Santa Catarina e Rio de Janeiro), seis com detecção média (São Paulo, Paraná, Distrito Federal, Minas Gerais, Rio Grande do Norte e Alagoas) enquanto que a maior parte dos estados brasileiros possuem detecção alta (Amazonas, Paraíba, Espírito Santo, Amapá, Bahia, Mato Grosso do Sul, Acre, Sergipe, Ceará e Goiás), muito alta (Pernambuco, Roraima Rondônia, Pará, Piauí e Maranhão) e hiperendêmica (Tocantins e Mato Grosso). O estado do Paraná é o quinto estado brasileiro com menor detecção a cada 100 mil habitantes (4,86) [BRASIL 2018].

<sup>1</sup>PPGMNE/UFPR. e-mail: *vsehaber@gmail.com*. Agradeço à Capes e ao CNPq pelo apoio financeiro.

<sup>3</sup>PPGMNE/DEST/UFPR. e-mail: *paulojus@ufpr.br*

<sup>4</sup>DEST/UFPR. e-mail: *eliaskr@ufpr.br*

Com o intuito de melhorar suas atividades de vigilância epidemiológica, a Secretaria de Saúde do Estado do Paraná (SESA) tem o interesse de estudar a evolução da hanseníase no Estado além de obter mapas de previsão 1 e 2 tempos à frente para auxiliar no seu monitoramento da doença. Consoante a este propósito, este trabalho utilizou uma modelagem espaço-temporal para explicar os casos novos anuais de hanseníase nos municípios do Estado do Paraná, no período entre 1980 e 2013.

## 2 Materiais e Métodos

O coeficiente de detecção anual de hanseníase é um indicador epidemiológico cuja finalidade é mostrar a tendência anual e medir a intensidade das atividades de detecção de casos novos da doença [BRASIL 2010]. Este coeficiente é dado pela razão

$$CD = \frac{\text{Número anual de casos novos diagnosticados}}{\text{População anual}} \cdot 100.000 \quad (1)$$

a fim de obter o número de casos novos residentes no município diagnosticados por ano a cada 100.000 habitantes. Os valores de referência utilizados para interpretar a intensidade da detecção são: baixa ( $CD < 2$ ), média ( $2 \leq CD < 10$ ), alta ( $10 \leq CD < 20$ ), muito alta ( $20 \leq CD < 40$ ) e hiperendêmica ( $CD \geq 40$ ) [BRASIL 2010].

O CD é definido de forma semelhante a taxa bruta, porém difere-se em termos práticos: o indivíduo pode não saber que possui hanseníase (período de incubação estimado de 5 anos) até a detecção da doença, que acontece por meio de exame clínico feito pelo profissional de saúde. Uma vez que os sintomas da hanseníase são detectados, é feito o registro no SINAN pela unidade de saúde.

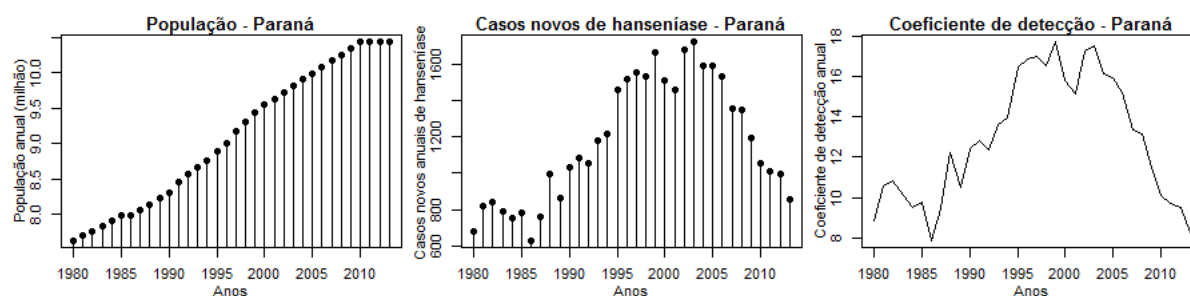


Figura 1: Gráficos de perfis: população do Paraná (à esquerda), casos novos anuais de hanseníase no Paraná (centro) e coeficiente de detecção (à direita)

A Figura 1 mostra o perfil do CD, tal como é indicado na Equação (1). É possível verificar que a tendência do CD é muito próxima a de casos novos de hanseníase. Note que, conforme a população do estado foi crescendo, os registros de casos novos de hanseníase também cresceram até 2003. Após, mesmo com a tendência de crescimento da população até 2010 (Censo), há um decaimento dos registros de casos novos anuais de hanseníase. Por causa disso, o CD apresenta um decaimento mais rápido a partir de 2005. A mudança da tendência é devida ao início do uso do medicamento de cura, chamado poliquimioterapia,

o qual passou a ser distribuído no estado a partir de 1994. E, por volta de 2000, todas as unidades de saúde dos municípios já prescreviam esse medicamento aos pacientes de hanseníase. Observe que, em 2013, o coeficiente de detecção anual do estado está abaixo de 10, o que é recomendável pela OMS. Porém, isso não ocorre de forma uniforme para todos os municípios do estado, conforme mostra a Figura 2. (Por conveniência, as populações de 2011, 2012 e 2013 foram assumidas constantes e iguais a população levantada pelo censo de 2010.)

Durante a exploração dos dados, foi verificado que o número de municípios aumentou no período de estudo. Em 1980, o estado do Paraná se apresentava com 290 municípios, chegando a 399 em 1997. E esta configuração do estado permanece a mesma até os dias atuais. Dentro do período de 1980 a 2013, a estrutura de dependência da vizinhança implicaria alguma complexidade espacial e, por conveniência, consideramos este conjunto de dados como geostatístico, onde as localizações são os centróides das áreas dos municípios [DIGGLE et. al. 2013].

Neste estudo, o número de casos novos anuais de hanseníase registrados por município paranaense foi considerado como variável dependente, para a qual se assumiu ter distribuição Poisson. Nesse sentido, a taxa dessa distribuição de probabilidades pode ser interpretada em termos do coeficiente de detecção anual (sem o fator de multiplicação 100.000). Logo, a especificação do modelo ajustado é dada por

$$\begin{aligned} \mathbf{Y}(\mathbf{x}_t, t) | \boldsymbol{\lambda}(\mathbf{x}_t, t) &\sim \text{Poisson}(\boldsymbol{\lambda}(\mathbf{x}_t, t) \mathbf{O}(\mathbf{x}_t, t)) \\ \log(\boldsymbol{\lambda}(\mathbf{x}_t, t)) &= \boldsymbol{\beta}_0 + \boldsymbol{\omega}(\mathbf{x}_t, t) \end{aligned} \quad (2)$$

onde  $\mathbf{Y}(\mathbf{x}_t, t)$  é o número de casos novos anuais de hanseníase por centróide de município  $\mathbf{x}_t$  no tempo  $t$ , com  $t = 1, \dots, 34$ ;  $\boldsymbol{\lambda}(\mathbf{x}_t, t)$  e  $\mathbf{O}(\mathbf{x}_t, t)$  são, respectivamente, a taxa de casos novos anuais de hanseníase e a população (*offset*) por centróide do município  $\mathbf{x}_t$  no tempo  $t$ ;  $\boldsymbol{\beta}_0$  é um efeito constante no espaço e no tempo;  $\boldsymbol{\omega}(\mathbf{x}_t, t)$  representa um efeito latente Gaussiano do processo espaço-temporal, o qual é estabelecido por meio de propriedade dinâmica autoregressiva de primeira ordem, com inovações espacialmente correlacionadas, conforme [BLANGIARDO 2015], o qual tem  $\sigma^2$  como variância espaço-temporal,  $\phi$  como parâmetro de dependência temporal e  $\rho$  como parâmetro de dependência espacial.

Neste trabalho, o *software* R, versão 3.4.3, foi utilizado [R Core Team 2018]. Mais especificamente, utilizamos a biblioteca INLA a qual disponibiliza o método de análise baseado em equações diferenciais parciais estocásticas (SPDE) para a análise espaço-temporal dos casos novos anuais de hanseníase no estado do Paraná, no período entre 1980 e 2013.

### 3 Resultados

Os resultados obtidos por meio do ajuste do modelo geostatístico espaço-temporal Poisson, usando SPDE/INLA, estão dispostas na Tabela 1.

Tabela 1: Estimativas do parâmetros do modelo Poisson

Estimativas	Média	E. Padrão	Quantil 0,025	Quantil 0,5	Quantil 0,975
$\hat{\beta}_0$	-9,2202	0,0683	-9,3543	-9,2202	-9,0862
$\hat{\sigma}$	0,7531	0,0217	0,711	0,7528	0,7965
$\hat{\phi}$	0,9269	0,0056	0,9154	0,9270	0,9374
$\hat{\rho}$	0,5082	0,0268	0,4573	0,5077	0,5625

Por meio da Tabela 1, foi verificado que os parâmetros estão contidos no intervalos de credibilidade de 95%. Foi estimado um efeito constante na taxa de 9,902. Além disso, foi estimada uma forte dependência temporal ( $\hat{\phi} = 0.927$ ) no modelo proposto, e um alcance prático de, aproximadamente, 50 km (correlação espacial fixada em 0,13). Uma animação dos mapas das médias a posteriori do modelo podem ser encontrada aqui (extrair). Esta é uma forma de acompanhar a evolução do CD no estado no período de 1980 a 2013.

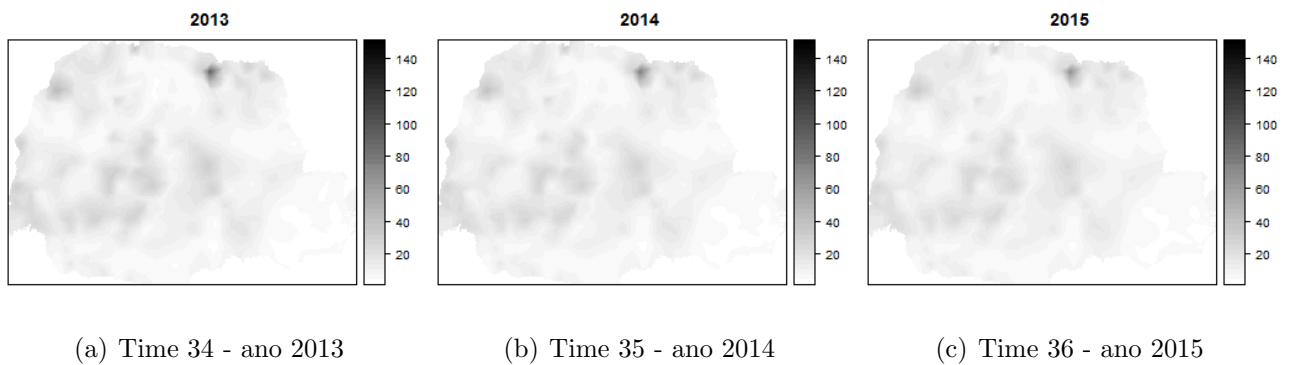


Figura 2: Média a posteriori da taxa multiplicada por 100.000, considerando o último ano de registro de dados (2013) e mapas de previsões 1 e 2 passos à frente (2014 e 2015)

De acordo com a Figura 2, no tempo 34 (ano de 2013), pode-se afirmar que não há regiões com CD muito alto e hiperendêmico, exceto por uma pequena área no topo do gráfico, a qual corresponde a área dos municípios de Rancho Alegre, Leópolis, Sertaneja, Uraí, Jataizinho e Ibiporã. Em geral, os CDs variam de baixo a médio sobre o estado do Paraná, porém a região sudeste tende a apresentar as menores taxas.

Os tempo 35 (ano 2014) e 36 (ano 2015) correspondem aos mapas de previsão espacial 1 e 2 passos à frente, respectivamente. Devido a forte dependência temporal, tais gráficos mostram que o padrão espacial demorará a tender a média do processo. E essa informação é bastante razoável uma vez que o período de latência da doença é longo, podendo variar de 6 meses a 5 anos [BRASIL 2009]. Com base nesses mapas, é possível verificar as regiões do estado que precisam de mais atenção.

## 4 Considerações Finais

Alguma covariáveis foram investigadas para auxiliar a explicação da variação no número de casos novos anuais de hanseníase, contudo muitas dessas informações estão incompletas



(registros a partir de 1995).

A análise espaço-temporal apresentada neste trabalho é uma ferramenta muito útil para a SESA e vem de encontro com a necessidade de monitoramento e vigilância da hanseníase. Da mesma forma, outras doenças podem ser monitoradas utilizando a metodologia de análise apresentada neste trabalho.

## Referências

- [BLANGIARDO 2015]BLANGIARDO, M.; CAMELETTI, M. *Spatial and Spatio-temporal Bayesian Models with R - INLA*.Chichester: Wiley, 2015.
- [BRASIL 2006]BRASIL, Ministério da Saúde. Secretaria de Vigilância em Saúde. Fundação Oswaldo Cruz. *Abordagens espaciais na saúde pública*. Série B. Textos Básicos de Saúde. Série Capacitação e Atualização em Geoprocessamento em Saúde. v. 1. Brasília, 2006.
- [BRASIL 2009]BRASIL, Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. *Guia de vigilância epidemiológica*. Série A. Normas e Manuais Técnicos. ed. 7. Brasília, 2009.
- [BRASIL 2010]BRASIL, Ministério da Saúde. *Roteiro para uso do sistema de informação de agravos de notificação - Sinan NET HANSENÍASE*. Brasília.
- [BRASIL 2018]BRASIL, Ministério da Saúde. *Hanseníase. Indentificou. Tratou. Curou*. <http://portalarquivos2.saude.gov.br/images/pdf/2018/janeiro/31/Campanha-do-Dia-Mundial-de-Luta-Contra-a-Hansen--ase-2018.pdf>
- [DIGGLE et. al. 2013]DIGGLE, P. J.; MORAGA, P.; ROWLINGSON, B.; TAYLOR, B. M. *Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm*. Statistical Science. a. 28 n. 4, p. 542-563, 2013.
- [IBGE 2018]BRASIL, Instituto Brasileiro de Geografia e Estatística (IBGE). <https://ww2.ibge.gov.br/home/disseminacao/eventos/missao/default.shtm>, 2018.
- [PARANÁ 2015]PARANÁ, Tribunal de Contas do Paraná (TCE-PR). *Estudo de Viabilidade Municipal*. <http://www1.tce.pr.gov.br/conteudo/estudo-de-viabilidade-municipal/296299/area/10>, 2015.
- [PENNA 2011]Penna, G. O. et. al. *Doenças dermatológicas de notificação compulsória no Brasil*, vol. 86 n. 5, p. 865-877, 2011.
- [R Core Team2018]R Core Team. *R: A language and environment for statistical computing*.R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [OMS 2017]OMS, Organização Mundial de Saúde. *Global leprosy update, 2016: accelerating reduction of disease burden*. Weekly Epidemiological Record. n. 92, v. 35, p. 501-519, Geneva, 2017.

## **Análise estatística do consumo de Metilfenidato nos estados de Minas Gerais, Rio de Janeiro e São Paulo entre os anos de 2008 a 2014**

**Rafaela da Silva Gomes<sup>1 3</sup>**

**Fábio Henrique de Jesus dos Santos<sup>2 3</sup>**

**Luciene Resende Gonçalves<sup>2 3</sup>**

O fenômeno da medicalização, hoje, pode ser pensado como resultado de um efeito conjugado que o discurso da ciência e o discurso capitalista impõem à civilização. O aumento de consumo de medicamentos psicofármacos, usados em larga escala para tratamento de distúrbios psicológicos, comportamentais e educacionais, observados principalmente entre crianças e adolescentes, traz efeito tranquilizador para os pais, fazendo com que as situações na escola e na família sejam momentaneamente controladas. No entanto, esses efeitos não são duradouros; pelo contrário, podem ser devastadores porque para a psicanálise o que fica silenciado pela química, retorna de forma violenta posteriormente. O metilfenidato é um princípio ativo de medicamentos psicofármacos como Ritalina® e Concerta® que vem apresentando um consumo excessivo ao longo dos anos. É utilizado como estimulante do sistema nervoso central, indicado como auxiliar de intervenções psicológicas, educacionais e sociais. O consumo do medicamento, inclusive não prescrito, pode causar dependência e abusos, por isso é necessária informação segura para promover seu uso racional. O objetivo é mostrar por meio de dados estatísticos utilizando técnicas de estatística básica e séries temporais, no *software R*, o comportamento do consumo de metilfenidato nos estados de Minas Gerais, São Paulo e Rio de Janeiro analisando os anos de 2008 a 2014. Os resultados esperados são contribuir para uma reflexão sobre o uso saudável de medicamentos no país, apontar possíveis distorções na utilização de metilfenidato e dar transparência aos dados apresentados para subsidiar novas pesquisas e políticas públicas que questionem e façam o enfrentamento à medicalização.

**Palavras-chave:** *Medicalização, Metilfenidato, Psicofármacos, Software R.*

---

<sup>1</sup> ICSA– Unifal. e-mail rafadasilva43@yahoo.com.br

<sup>2</sup> ICSA– Unifal.

<sup>3</sup> Agradecimento a FAPEMIG pelo apoio financeiro.

## **Análise temporal do número de mortes por pneumonia em Minas Gerais**

**Lucas Roberto de Castro<sup>13</sup>**

**Emilly Malveira de Lima<sup>23</sup>**

Doenças respiratórias afetam crianças, jovens e adultos e, de acordo com a Organização Mundial da Saúde (OMS), é uma das principais causadoras de mortalidade no mundo. Além disso, a *World Health Statistics* (2012) destacou que doenças do trato respiratório inferior, incluindo a pneumonia, constam entre as três principais causas de potenciais anos perdidos. A pneumonia é uma doença caracterizada pela inflamação dos pulmões causada por agentes infecciosos tais como fungos, vírus e bactérias. Os sintomas incluem tosse com catarro ou pus, febre, calafrios e dificuldade respiratória. Por ser uma doença muito comum no Brasil, a pneumonia torna-se um grande problema de saúde pública. Neste contexto, com o intuito de investigar a evolução temporal e detectar padrões nos períodos em que ocorre maior mortalidade causada por esta doença no estado de Minas Gerais, foi aplicado um modelo linear generalizado dinâmico Poisson em dados obtidos a partir levantamento do número de mortes mensais causadas por pneumonia no estado, ocorridas entre os anos de 2011 a 2015 e registradas no Sistema de Informação sobre Mortalidade mantido pelo Ministério da Saúde (SIM-MS). A partir dos resultados alcançados, constatou-se que o número de óbitos no estado passou por um período de crescimento entre os anos de 2012 a 2013 e que se mantém em um nível praticamente constante desde então, além disso, foi observado que o maior número de mortes ocorre no período de março a abril, que marca o início da primavera na região caracterizada por temperaturas mais amenas.

**Palavras-chave:** *Modelo linear generalizado dinâmico, Sazonalidade, Séries temporais.*

---

1 Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, UFLA. lrcastro@posgrad.ufla.br

2 Programa de Pós-Graduação em Estatística, UFMG.

3 Agradecimento à FAPEMIG e CNPq pelo apoio financeiro.

# Análise temporal dos casos de Diabetes Mellitus na Região de Saúde de Feira de Santana no período de 1998 a 2016

Raiara dos Santos Pereira Dias <sup>1 3</sup>

Aloísio Machado da Silva Filho <sup>2 3</sup>

Edna Maria de Araújo <sup>2 3</sup>

Everaldo Freitas Guedes <sup>4 3</sup>

O presente estudo tem como objetivo analisar as taxas mensais de internação por Diabetes Mellitus na região de saúde de Feira de Santana-Ba no período de 1998 a 2016, tendo como principal método de análise o Detrended Fluctuation Analysis-DFA. **Dados:** A região de saúde de Feira de Santana é composta atualmente de 28 municípios (6,7 % dos municípios baianos). Foram utilizados os dados de população do IBGE e a base de dados disponibilizada pelo DATASUS. **Metodologia:** O teste estatístico do DFA alicerçou-se no método desenvolvido por Kristoufek (2010), que consiste na simulação de 10.000 séries aleatórias independentes e identicamente distribuídas (*iid*) de um processo gaussiano  $N(0, 1)$  de tamanho igual a  $N$  (Tamanho da série objeto de estudo). O referido teste foi realizado devido ao fato do DFA apresentar resultados robustos a partir de séries com  $N$  igual a pelo 1.000 pontos ( $N \geq 1.000$ )(CARBONE, et al, 2004). **Resultados:** A análise descritiva da série histórica revelou uma taxa média de 4,4 casos por 100 mil habitantes com uma variação relativa em torno da média de 53,9% e assimetria positiva denotando concentração de taxas abaixo da média e variação moderada. O DFA identificou e mensurou dependência temporal persistente, ou seja expoente alfa maior que 0,50 ( $\alpha > 0,50$ ), porém sem significância estatística, ao nível de 99% de confiança. **Considerações finais:** Espera-se que os resultados e modelagem estatística executada nesta pesquisa motivem pesquisadores, gestores públicos e privados interessados no tema a estudar demais regiões de saúde do Estado da Bahia e sirva como mais uma metodologia destinada ao monitoramento dos casos de Diabetes Mellitus.

**Palavras-chave:** *Diabetes Mellitus, Série temporal, Autocorrelação.*

---

<sup>1</sup>Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente da UEFS. e-mail: [raiara25@gmail.com](mailto:raiara25@gmail.com)

<sup>2</sup>Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente da UEFS.

<sup>3</sup>Agradecimento ao CNPq e a FAPESB pelo apoio financeiro.

<sup>4</sup>Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial do SENAI CIMATEC.

## Análise temporal dos homicídios registrados na Bahia no período de 2012 a 2016: Uma abordagem com o modelo ARIMA

Sátira Izabel Soares Nunes <sup>1 3</sup>

Aloísio Machado da Silva Filho <sup>2 3</sup>

Carlos Alberto Lima da Silva <sup>2 3</sup>

Este estudo analisou as taxas mensais de homicídios dolosos que foram registrados diariamente no Estado da Bahia, no período de janeiro de 2012 a dezembro de 2016, com objetivo de modelar a série temporal através dos modelos auto-regressivos de médias móveis baseado na metodologia de *George Box* e *Gwilym Jenkins*. Os dados referentes ao número absoluto dos homicídios dolosos foram coletados no site da Secretaria de Segurança Pública do Estado da Bahia e calculadas as respectivas taxas com os dados populacionais fornecidos pelo Instituto Brasileiro de Geografia e Estatística. A série analisada não apresentou tendência nem sazonalidade nos dados mensais, a taxa mais elevada no período ocorreu no mês de fevereiro de 2012 (4,3 por 100 mil habitantes), e a mínima em junho e em setembro de 2014 (2,5 por 100 mil habitantes). O modelo mais adequado, segundo a metodologia adotada nesta pesquisa, foi, ARIMA(8,1,1) por ter atendido os pressupostos na análise dos resíduos e menor AIC. Foi realizada a previsão para seis meses subsequentes ao período analisado e o modelo ARIMA(8,1,1) apresentou menor resíduo entre o valor predito e o valor observado.

**Palavras-chave:** *Séries temporais, Homicídios, ARIMA.*

---

<sup>1</sup>UNIJORGE. e-mail: *satira.ufba@gmail.com*

<sup>2</sup>Universidade Estadual de Feira de Santana - UEFS.

<sup>3</sup>Agradecimento a Universidade Estadual de Feira de Santana e a Secretaria de Segurança Pública do Estado da Bahia.

## **Avaliação da qualidade de vida de indivíduos com disfunções temporomandibulares em uma instituição de ensino superior: uma aplicação da teoria de resposta ao item**

Juscelia Dias Mendonça<sup>1</sup>, Nadja Gomes Machado<sup>2</sup>, Nívio Lotufo-Neto<sup>3</sup>, Gilmar Jorge de Oliveira Junior<sup>4</sup>, Danielle Gonçalves de Oliveira Prado<sup>5</sup>

### **RESUMO**

Com o questionário SF-36 é possível verificar e classificar os indivíduos em relação à qualidade de vida utilizando. Neste trabalho avaliou-se a qualidade de vida de indivíduos com disfunções temporomandibulares em uma instituição de ensino superior, aplicando o modelo de resposta gradual da teoria de resposta ao item. Os itens 1, 20, 26, 27, 21, 36, 23, 30 e 34 do questionário apresentaram discriminação negativa. Portanto, podemos concluir que são itens que apresentam problemas e devem ser revistos no questionário SF-36.

<sup>1</sup> Professora Adjunta da UFMT

<sup>2</sup> Professora Adjunta no IFMT

<sup>3</sup> Docente da Graduação em Estatística na UFMT

<sup>4</sup> Professor Assistente da UFMT

<sup>5</sup> Professora Adjunta da UFTPR

## **Avaliação de possíveis causas de lesões musculoesqueléticas em policiais militares por meio de modelos logísticos**

**Patricia Mendes dos Santos<sup>1</sup>**

**Ana Cláudia Vasconcelos da Costa<sup>2</sup>**

**Izabela Regina Cardoso de Oliveira<sup>3</sup>**

**Giancarla Aparecida Botelho Santos<sup>4</sup>**

**Marcelo Ângelo Cirillo<sup>3</sup>**

A segurança pública tem merecido destaque devido a sua importante atuação na sociedade, uma vez que a criminalidade vem ganhando espaço cada vez maior. Tal fato tem exigido dos policiais militares maiores cargas de trabalho, aumento do estresse e menor tempo disponível para um bom preparo físico. A falta de preparo tem resultado em aumento de peso e alta incidência de lesões musculoesqueléticas nos policiais militares. Sendo assim a avaliação dos possíveis fatores causadores destas lesões é fundamental, uma vez que o efetivo militar é baixo e o afastamento pode causar prejuízos à corporação e à sociedade. O presente trabalho tem como objetivo identificar as possíveis causas de lesões osteomusculares em policiais militares e verificar se há relação entre a ocorrência dessas lesões e a prática de atividade física. Os dados foram obtidos por meio de um questionário respondido por 120 policiais militares do 8º batalhão, na região de Lavras, MG. A pesquisa foi realizada em 2014 e a participação foi voluntária. Modelos logísticos foram utilizados para avaliar os efeitos do uso do cinto de armamento, do colete e do perfil do entrevistado em relação à atividade física (sedentário ou ativo) na ocorrência ou não de dor osteomuscular. Todos os modelos foram ajustados no programa R. Dentre os resultados, destaca-se que as chances de um indivíduo relatar dor osteomuscular são quatro vezes para indivíduos sedentários em relação aos ativos. Esse resultado demonstra a importância de programas para conscientização, apoio e incentivo da prática de atividade física por esses profissionais.

**Palavras-chave:** *Atividade física, Lesões osteomusculares, Razão de chances.*

---

<sup>1</sup> Doutoranda em Estatística e Exp. Agropecuária, Universidade Federal de Lavras - UFLA. e-mail: patymendesdossantos@hotmail.com

<sup>2</sup> Graduanda em Educação Física, UFLA.

<sup>3</sup> Professor do Departamento de Estatística, UFLA.

<sup>4</sup> Professora do Departamento de Ciências da Saúde UFLA.

\* Os autores agradecem à FAPEMIG e CAPES pelo apoio financeiro.

## **Compatibilização das Bases do SIM e SINASC: aplicação do método linkage para o município de João Pessoa-PB, 2011 - 2015**

### **RESUMO**

Avanços na qualidade das informações das estatísticas vitais ocorreram ao longo dos anos no Brasil, contudo regiões como Norte e Nordeste ainda possuem dados deficientes, prejudicando o planejamento de saúde, cujos dados deles dependem. Estudos de compatibilidade de fontes de informações sobre estas estatísticas para o município de João Pessoa/PB são inexistentes. Este estudo teve como objetivo avaliar a qualidade da cobertura, completude e fidedignidade de informações dos sistemas SIM e SINASC sobre a mortalidade infantil no município de João Pessoa de 2011 a 2015. Através de um estudo ecológico, avaliou-se a qualidade das informações, através do percentual de preenchimento, da taxa de vinculação das bases após pareamento pelo método do *linkage* determinístico, e a concordância pelo Índice Kappa. Considerou-se o código único (número da declaração de nascido vivo) para emparelhamento da declaração de óbito com sua respectiva declaração de nascido vivo. A taxa de vinculação só foi completa no primeiro ano da implantação da nova DO, em 2011, reduzindo-se para 68,8% em 2015. Apesar do avanço importante na qualidade do preenchimento das variáveis da DN e DO, a taxa de vinculação foi considerada apenas satisfatória. Isto significa que ainda há espaço para melhoras na precisão entre as fontes oficiais de dados utilizadas. Sendo assim, as secretarias municipais de saúde do país poderiam se apropriar desta metodologia de forma sistemática e contínua para aperfeiçoar a qualidade das estatísticas vitais.

**Palavras-chave:** *Mortalidade infantil, Linkage, Qualidade de dados.*



## Descobrimos as cidades mais influentes para os casos de dengue no Paraná usando grafos

Mônica Marin de Souza <sup>1</sup>

Haroldo Valentin Ribeiro <sup>2</sup>

Edson Zangiacomi Martinez <sup>3</sup>

Em 2007, pela primeira vez na história, a população mundial urbana superou a rural. Se esse processo de urbanização persistir no ritmo atual, 2/3 da população mundial será urbana até o ano de 2050. A vida urbana produziu muitos benefícios para a humanidade, mas também conduziu a muitos problemas como poluição e epidemias. O aumento da incidência de doenças epidêmicas reflete, em grande parte, a maior mobilidade de pessoas através das redes de transporte e o crescimento no número de pessoas vivendo em áreas de grande densidade populacional. No Brasil, uma das endemias que causa grandes prejuízos e preocupações é a dengue. Em 2009, quando a endemia teve seu pico, mais de 1 milhão de brasileiros foram infectados pelo vírus da dengue. Além disso, o Brasil é o país que mais reportou casos de dengue no presente século. Por conta disso, de sua dimensão territorial e do grande número de municípios, o Brasil é ideal para investigar padrões epidêmicos associados à dengue. Nesse trabalho, utilizamos medidas de correlação e teoria de grafos para estabelecer relações entre a evolução dos casos de dengue nas cidades paranaenses que apresentaram ao menos um caso da doença por ano ao longo do período de 2006 a 2012. Após construir um grafo representando as relações entre as cidades do Paraná, observamos quais cidades têm maior e menor influência, de acordo com a centralidade do grau, para a evolução dos casos de dengue no estado.

**Palavras-chave:** *epidemia, dengue, cidades, teoria de grafos.*

---

<sup>1</sup>PBE - UEM. e-mail: [monicamarins@gmail.com](mailto:monicamarins@gmail.com)

<sup>2</sup>DFI - UEM.

<sup>3</sup>FMRP - USP.

## Estratégias de modelagem estatística de pressão arterial sanguínea: uma aplicação ao estudo de hipotensão pós-exercício

Denize Palmito dos Santos<sup>1 5</sup>

Izabela Regina Cardoso de Oliveira<sup>2 5</sup>

Lucas Roberto de Castro<sup>3 5</sup>

Giancarla Aparecida Botelho Santos<sup>4 5</sup>

Em Ciências da Saúde os estudos controlados desempenham um papel importante no desenvolvimento de técnicas que ajudam na prevenção ou na cura de doenças. Frequentemente, nessa área utilizam-se medidas repetidas no tempo, pois as análises estatísticas aplicadas a dados dessa natureza permitem vantajosamente observar as possíveis mudanças na variável resposta. A metodologia adotada para a análise de dados longitudinais desenvolve o uso de modelos mistos, em que o método ML ou REML podem ser usados na estimação dos parâmetros. A hipertensão é um problema mundial de saúde que, se não tratada corretamente, pode levar o hipertenso a sofrer consequências graves. A pressão arterial (PA) é aferida em termos da pressão arterial sistólica (PAS) e diastólica (PAD) e, através de cálculos simples, obtém-se a pressão arterial média (PAM). Atualmente buscam-se métodos não farmacológicos, como a prática de atividades físicas, para o controle da PA. Um estudo desenvolvido em Lavras, MG, avaliou a PA de 26 indivíduos normotensos, de ambos os sexos, submetidos a dois tipos de treinamentos (resistência e hipertrofia). Os valores de PA foram obtidos ao longo de 18 sessões. Para avaliar o efeito (fixo) do tipo de exercício na redução da PA, foram consideradas e comparadas duas alternativas de modelagem (mista): modelagem conjunta de PAS e PAD e modelagem da PAM. Em todos os casos, sexo, idade e medida inicial da PA foram usados como covariáveis. Verificou-se que ambos os tipos de exercícios físicos diferem em relação ao controle da PA. Todas as análises foram realizadas no programa R.

**Palavras-chave:** *Dados longitudinais, Efeitos aleatórios, Modelos Mistos, Respostas bivariadas.*

---

<sup>1</sup>Mestranda em Estatística e Experimentação Agropecuária. DES - UFPA. e-mail: [denizepalmito@gmail.com](mailto:denizepalmito@gmail.com)

<sup>2</sup>DES - UFPA. E-mail: [izabela.oliveira@des.ufpa.br](mailto:izabela.oliveira@des.ufpa.br)

<sup>3</sup>DES - UFPA. E-mail: [lucasroberto.castro@gmail.com](mailto:lucasroberto.castro@gmail.com)

<sup>4</sup>DSA - UFPA. E-mail: [giancarla.ufpa@gmail.com](mailto:giancarla.ufpa@gmail.com)

<sup>5</sup>Agradecimento ao CNPq e FAPEMIG pelo apoio financeiro.

**Resumo:** Este trabalho tem o objetivo de relacionar os casos de internações por gripe e outras doenças graves das vias aéreas superiores e o clima das regiões do Brasil. Foi analisada uma amostra de 1812 cidades retiradas no banco de dados DATASUS dos meses de maio a agosto de 2010 utilizando o método de Krigagem Ordinária e Krigagem simples para interpolação e geração dos mapas.

## Estudo de casos de dengue na cidade de Dourados/MS utilizando séries temporais

Alessandra Querino da Silva<sup>1 4</sup>

Hagar Gonçalves Borges<sup>2 4</sup>

Luciano Antonio de Oliveira<sup>1 3 4 5</sup>

Thelma Sáfiadi<sup>3 5</sup>

A dengue é apontada como um dos maiores problemas de saúde pública mundial. No Brasil a incidência de dengue é registrada desde o início dos anos de 1980 trazendo transtornos principalmente as populações mais vulneráveis ao avanço da doença. O presente trabalho teve por objetivo estudar número de casos de dengue mensais de 2009 a 2016 na cidade de Dourados/MS. Aplicou-se a metodologia de séries temporais utilizando modelos de Box e Jenkins, que possibilitou incorporar um componente sazonal, útil em situações em que a série temporal tem picos e declínios periódicos recursivos ao longo de um período. Para a seleção do melhor modelo utilizou-se o critério de Akaike (AIC) e as análises foram realizadas com recursos do software R. Observou-se um grande surto de dengue na cidade no ano de 2010, uma sazonalidade de 12 meses para os picos do número de casos de dengue e o melhor modelo que se ajustou aos dados foi SARIMA (2,1,3)(1,1,1)<sub>12</sub>.

**Palavras-chave:** *Dengue, Dourados/MS, Séries Temporais.*

---

<sup>1</sup> FACET – UFGD e-mail: alessandrasilva@ufgd.edu.br

<sup>2</sup> FAEN – UFGD

<sup>3</sup> DES – UFLA

<sup>4</sup> Agradecimento a FUNDECT e ao CNPq pelo apoio financeiro.

<sup>5</sup> Agradecimento a FAPEMIG pelo apoio financeiro.

## **Fatores de risco para o baixo e elevado peso ao nascer com base nas Declarações de Nascidos Vivos**

**Carlos Sérgio Araújo dos Santos<sup>1</sup>**

**Tiê Dias de Farias<sup>2</sup>**

**Neir Antunes Paes<sup>3</sup>**

Este estudo teve como objetivo identificar fatores de risco associados ao baixo e elevado peso ao nascer das crianças de João Pessoa – PB registradas em 2015. Realizou-se um estudo transversal, no qual foram utilizadas 6085 Declarações de Nascidos Vivos. Utilizou-se a regressão logística ordinal para verificação da associação. A variável dependente foi o peso de nascimento e as independentes foram relacionadas às características maternas, do parto e do recém-nascido. Os determinantes do peso ao nascer significativos ao nível de 1% foram: duração da gestação, tipo de gravidez, tipo de parto, número de consultas de pré-natal, sexo, índice Apgar no quinto minuto e escolaridade da mãe. Os determinantes do baixo e elevado peso ao nascer permite aos profissionais de saúde elaborar políticas públicas das áreas de saúde materno-infantil, minimizando o risco de agravos à saúde.

**Palavras-chave:** *Fatores de risco, Peso ao nascer, Recém-nascido.*

---

1 PPG em Modelos de Decisão e Saúde – UFPB. e-mail: carlossergioaraujo@gmail.com

2 PPG em Modelos de Decisão e Saúde – UFPB.

3 Departamento de Estatística – UFPB.

## Joint Model para um estudo de coorte de HIV coinfectados por Hepatites B e C

Beatriz Regina Brum <sup>1 3</sup>

Omar Neves Pereira <sup>2</sup>

Claudinei Mesquita Silva <sup>3</sup>

Isolde Terezinha Previdelli <sup>2</sup>

Estudos que têm em vista determinar associações entre respostas longitudinais e de sobrevivência são comumente analisados de forma separada, com o avanço teórico estatístico e computacional, permite-se analisar conjuntamente ambos os processos, possibilitando obter estimativas mais eficientes no processo longitudinal e de menor viés no modelo de sobrevivência. E ainda, os modelos conjuntos têm grande apelo de interpretação biológica, pois quantifica o efeito do biomarcador na função de risco. Propõem-se nesta pesquisa abordar um modelo conjunto, assumindo para a resposta longitudinal um modelo misto normal multivariado e para a sobrevivência o modelo de riscos proporcionais. Neste estudo, utilizou-se dados de uma coorte retrospectiva obtida no estado do Paraná/Brasil, entre os anos de 2002 a 2014, de indivíduos portadores da Imunodeficiência Humana (HIV), coinfectados por hepatites e expostos à terapia antirretroviral combinada (CART). A pesquisa foi composta por três grupos: controle (HIV), infecção por hepatites B e C. Modelou-se a distribuição conjunta do tempo de início da CART até atingir a razão mínima considerada entre linfócitos T (LT) auxiliares (CD4), e citotóxicos (CD8) e a contagem do biomarcador LT CD4, em relação a essa razão. O Joint Model ajustado, sob aspecto longitudinal apontou maior severidade para os pacientes HIV coinfectados pela hepatite C e, do ponto de vista da sobrevivência, os pacientes com HIV coinfectados por hepatite B indicam pior índice prognóstico, isso quando comparado ao grupo controle. Diante disso, o *Joint Model* mostrou-se adequado para se analisar conjuntamente dois processos, e mais que isso, ele corrobora com algumas evidências clínicas.

**Palavras-chave:** *Biomarcador, Joint Models, HIV, infecção por hepatites B e C.*

---

<sup>1</sup>PBE - UEM. e-mail: [beatrizbrum2009@hotmail.com](mailto:beatrizbrum2009@hotmail.com)

<sup>2</sup>PBE - UEM.

<sup>3</sup>PCS - UEM.

# Modelo de Poisson $k$ -Modificado: Uma Aplicação com Notificações de Óbitos Fetais nos Municípios do Estado da Bahia

Milene Alves Garcia <sup>1 4</sup>

Vitor Millani Negrão Bonini <sup>2</sup>

Katiane Silva Conceição <sup>3 4</sup>

No presente trabalho uma extensão da distribuição de Poisson foi considerada para modelar dados de contagem que apresenta alguma discrepância (alta, baixa ou até mesmo regular) na frequência da observação  $k$ , denominada distribuição Poisson  $k$ -Modificada. Essa nova distribuição tem como vantagem a possibilidade de modificar a probabilidade de ocorrência no ponto  $k$ , de tal forma que esta probabilidade pode ser maior ou menor comparada com a distribuição Poisson usual. Essa modificação é feita com a inclusão de um parâmetro adicional, denominado parâmetro de modificação. Em um contexto mais amplo, foi introduzido o modelo de regressão Poisson  $k$ -modificado pois, muitas vezes, há interesse em explicar o conjunto de dados de contagem utilizando-se variáveis exógenas (ou explicativas). Como ilustração, um conjunto de dados recente referente a notificações de óbitos fetais (ocorridos e registrados no lugar de residência da mãe) das cidades do Estado da Bahia em 2014 foi considerado.

**Palavras-chave:** *Dados de Contagem; Distribuição de Poisson;  $k$ -Deflação;  $k$ -Inflação.*

---

<sup>1</sup>PIPGEs - USP/UFSCar.

<sup>2</sup>ICMC - USP.

<sup>3</sup>ICMC - USP. E-mail: [katiane@icmc.usp.br](mailto:katiane@icmc.usp.br)

<sup>4</sup>Agradecimentos a FAPESP e CAPES pelo apoio financeiro.

# Modelo de Poisson ZMPS-GARMA Bayesiano: uma aplicação em dados de mortalidade

Aimée Shirozono <sup>1 4</sup>

Marinho G. Andrade <sup>2 4</sup>

Katiane S. Conceição <sup>3 4</sup>

As taxas de mortalidade e as causas são indicadores importantes para ter um controle sobre temas como saúde da população, desenvolvimento econômico, segurança e expectativa de vida. Dados sobre a mortalidade brasileira são categorizados para tornar-se claro as decorrências ocasionadas pelas mortes. Uma das categorias são de doenças relacionadas ao sistema auricular que são menos comuns de apresentarem sintomas muito graves que possam gerar um óbito. Nosso objetivo neste trabalho é modelar o número de mortes causadas por doenças de ouvido e da apófise mastoide no município de São Paulo (SP), de 1996 a 2015.

Este tipo de informação pode ser classificado como dados de contagem correlacionados no tempo. Devido à quantidade de zeros presentes na série, iremos propor o modelo Zero-Modificados série de potência (ZMPS), aplicado no contexto de séries temporais, utilizando os resultados dos modelos Generalizados Autoregressivos e de Médias Móveis (GARMA).

As distribuições ZMPS incluem fatores de inflação/déficit de zeros, já os modelos GARMA apresentam a utilização de um preditor linear ARMA para modelos generalizados. Com a ligação entre estas duas metodologias é possível trabalhar com séries temporais em dados de contagem, aplicando a modificação de zeros ao longo do tempo. Resultados com esta proposta vão ser apresentados para a distribuição de Poisson utilizando a inferência Bayesiana. Um estudo de simulação também será discutido para verificar a aderência do modelo proposto.

**Palavras-chave:** *Mortalidade, Poisson Zero-Modificado, Modelos GARMA.*

---

<sup>1</sup>PIPGEs - USP/UFSCar.

<sup>2</sup>ICMC - USP. e-mail: [marinho@icmc.usp.br](mailto:marinho@icmc.usp.br)

<sup>3</sup>ICMC - USP.

<sup>4</sup>Agradecimento a FAPESP e CAPES pelo apoio financeiro.



Modelo Espacial Bayesiano para a Distribuição Espacial da Ocorrência de  
Leishmaniose Visceral nos Estados de São Paulo, Minas Gerais, Rio de Janeiro e Paraná  
no ano de 2014

Jacqueline Domingues, José Silvio Govone

A leishmaniose visceral (LV) vem se expandindo para áreas urbanas de médio e grande porte e se tornou crescente problema de saúde pública no país, sendo uma endemia em franca expansão geográfica. Neste trabalho analisou-se os casos notificados no ano de 2014 da doença em humanos para os estados de São Paulo, Minas Gerais, Rio de Janeiro e Paraná. A região de estudo foi dividida por microrregiões de acordo com o IBGE. No ano em análise foram registrados 582 casos, para estes casos foi assumida uma distribuição de Poisson com parâmetro  $(E_k\theta_k)$ , em que  $E_k$  é o valor esperado na microrregião  $k$  e  $\theta_k$  o risco de LV na microrregião  $k$ . O modelo ajustado é dado por  $\ln(\theta_k) = \beta + \varphi_k$ . Para o efeito aleatório  $\varphi_k$  foi assumido uma distribuição CAR proposta por Leroux, Lei e Braslow (2000) com intuito de capturar a dependência espacial existente entre microrregiões vizinhas. Por fim, o modelo espacial Bayesiano indica quais as microrregiões de maior risco de LV.

# Modelos de Fração de Cura Aplicados aos Tempos de Sobrevivência de Pacientes Submetidos à Ligadura Elástica de Varizes no Esôfago

Agda Jéssica de Freitas Galletti <sup>1 2</sup>

Rogério Antonio Oliveira <sup>2</sup>

Giovanni Faria Silva <sup>3</sup>

Liciana Vaz de Arruda Silveira <sup>2</sup>

A cirrose é uma doença hepática assintomática que muitas vezes é descoberta quando o quadro é irreversível. A principal consequência da cirrose é o aumento da pressão na veia portal, que por sua vez, acarreta no surgimento de varizes e no seu respectivo rompimento, podendo ser fatal. Estudos relacionados à esta doença são muito importantes e um método estatístico bastante explorado nas ciências biomédicas é a análise de sobrevivência, que consiste em descrever o tempo até a ocorrência de um evento de interesse. No entanto, existem situações em que uma proporção da amostra não vivencia o desfecho de interesse, mesmo que acompanhado por um período longo de tempo. Nestes casos, tais observações são ditas imunes e as metodologias tradicionais de análise de sobrevivência não são indicadas. Logo, os modelos de fração de cura ou de longa duração são os utilizados nestas situações. Neste trabalho diversos modelos foram considerados para analisar os tempos de vida de pacientes submetidos à Ligadura Elástica de Varizes Esofágicas, ao qual foram anotados os tempos até o óbito durante o acompanhamento de 129 pacientes do Hospital das Clínicas da Faculdade de Medicina da UNESP, Campus Botucatu - SP, no período de 2006 a 2010. O modelo Weibull Modificado com fração de cura foi considerado adequado, indicando que pacientes mais velhos, homens, com alto grau da doença hepática e que não utilizaram o medicamento  $\beta$ -Bloqueador têm menor chance de serem sobreviventes de longa duração.

**Palavras-chave:** *Análise de sobrevivência, Ligadura Elástica, Modelos de Longa duração, Varizes no Esôfago, Weibull Modificada.*

---

<sup>1</sup>Agradecimento a CAPES pelo apoio financeiro. E-mail: agda.galletti@ibb.unesp.br

<sup>2</sup>Universidade Estadual Paulista (Unesp), Instituto de Biociências, Botucatu.

<sup>3</sup>Universidade Estadual Paulista (Unesp), Faculdade de Medicina, Botucatu.

# Modelos de Regressão de Poisson Zero-Modificado para Dados Longitudinais

Gabriela Cintra Raquel <sup>1 4</sup>

Marinho G. Andrade <sup>2 4</sup>

Katiane S. Conceição <sup>3 4</sup>

Este trabalho apresenta os modelos Zero-Modificados (e.g. Inflacionados ou Deflacionados de Zero) para representar dados de contagem longitudinais. Os modelos desenvolvidos neste trabalho tem como origem alguns dos modelos Zero-Modificados da família Série de Potência (Poisson, Binomial, Geométrico, Poisson Generalizado, COM-Poisson e Binomial negativo). No presente trabalho vamos considerar o modelo de regressão de Poisson Zero-Modificado no contexto de dados longitudinais para acomodar dados de contagem com excesso (ou déficit) de zeros e correlacionados no tempo. Serão considerados métodos de inferência baseados na abordagem clássica de verossimilhança. Técnicas de bootstrap serão consideradas no procedimento de inferência clássica. O trabalho também apresenta uma aplicação com o modelo de regressão de Poisson Zero-Modificado com efeito aleatório Normal para dados longitudinais para analisar um conjunto de dados referente a um ensaio clínico realizado com 59 pacientes epiléticos ao longo de 16 semanas, onde existe uma correlação entre o número de convulsões observadas no mesmo indivíduo. O ajuste do modelo com a inclusão do efeito aleatório Normal é responsável por captar a correlação entre os dados e eliminar a variabilidade intrínseca de cada indivíduo.

**Palavras-chave:** *Distribuição Série de Potência; Zero-Inflação; Zero-Deflação; Dados Longitudinais; Modelos misto.*

---

<sup>1</sup>PIPGEs - USP/UFSCar.

<sup>2</sup>ICMC - USP. e-mail: [marinho@icmc.usp.br](mailto:marinho@icmc.usp.br)

<sup>3</sup>ICMC - USP.

<sup>4</sup>Agradecimento a FAPESP e CAPES pelo apoio financeiro.

# Modelos Lineares Generalizados para dados de contagem com estrutura de autocorrelação temporal

Natacha Neves <sup>1 3</sup>

Ricardo Alves de Olinda <sup>1 3</sup>

Tiago Almeida de Oliveira <sup>1 3</sup>

Ednário Barbosa de Mendonça <sup>1 3</sup>

Modelos de séries temporais para dados de contagem são registros de frequência relativa da ocorrência de determinados eventos em sucessivos intervalos de tempo, e tem como característica importante a dependência temporal entre observações. As distribuições Poisson e binomial, pertencentes à classe dos Modelos Lineares Generalizados (MLG), são as mais conhecidas, e as mais utilizadas para modelar dados de contagem, no entanto sempre que existe sobredispersão, torna-se necessário recorrer a outras distribuições, nomeadamente à distribuição binomial negativa. Uma opção para melhorar o ajuste, visto que o MLG não é capaz de capturar a dependência ao longo do tempo, seria o uso de modelos de séries temporais para dados de contagem. Diante do exposto este trabalho teve por objetivo estudar as classes de modelos GARMA e INGARCH, que podem ser utilizados para modelar séries temporais de contagem que apresentam dependência temporal, a saber, com distribuições condicionais de Poisson, binomial e binomial negativa. Diante do exposto este trabalho vem contribuindo em pesquisas da UEPB, com aplicações na área da Saúde Pública, mostrando a importância de se utilizar novas metodologias capazes de englobar modelos lineares generalizados com estrutura de média móvel autorregressiva. Este estudo possibilitou a análise da relação entre o número de casos notificados de Doenças Diarreicas Agudas em crianças inseridas na faixa etária entre zero a nove anos de idade no município de Campina Grande e o índice pluviométrico do Açude Eptácio Pessoa, que abastece a cidade.

**Palavras-chave:** *Dados de contagem, Modelos autorregressivos, Sobredispersão.*

---

<sup>1</sup>Departamento de Estatística - UEPB. e-mail: [natacha\\_neves@hotmail.com](mailto:natacha_neves@hotmail.com)

<sup>3</sup>Agradecimento à Universidade Estadual da Paraíba pelo apoio financeiro.

## O problema da unidade de área modificável na análise da distribuição espacial da mortalidade infantil em Minas Gerais.

Walef Machado de Mendonça<sup>1 3</sup>

Taylor Oliveira Fidelis<sup>2 3</sup>

Diana Chaukat Chaib<sup>2 3</sup>

Quando uma estatística é calculada sobre duas unidades de área espaciais é comum que seja gerado diferenças em relação ao recorte geográfico. Segundo Haining (2003) essa é a característica que descreve o fenômeno do MAUP (*modifiable areal unit problem*). Este trabalho busca apresentar esse problema na distribuição espacial da mortalidade infantil no estado de Minas Gerais. O problema da unidade de área modificável surge da necessidade de se agregarem informações espaciais. Essa agregação pode ser necessária porque não se têm dados individuais, ou porque o próprio objeto de análise são unidades agregadas. Haining (2003) aponta dois tipos de MAUP: o problema de escala decorre dos diferentes níveis de agregação das unidades espaciais; o problema de partição decorre do próprio recorte das unidades (mantido o nível constante). Nesse estudo apresentamos o problema da partição. O maior nível de agregação utilizado foi o estado de Minas Gerais que foi dividido nas seguintes partições: municípios, microrregiões e mesorregiões. Foi feita a análise exploratória de dados espaciais para descrever os padrões de associação espacial através do coeficiente de autocorrelação I de Moran para cada um dos recortes mantendo-se o nível de agregação constante. Foi possível observar que quando computado sobre diferentes partições obtém-se resultados diferentes para o I de Moran. Quando calculado utilizando os dados dos municípios o I de Moran foi 0.453, com os dados das microrregiões 0.839 e quando utilizamos os dados de mesorregiões 0.524. Todos os testes foram significativos, indicando que há a existência de dependência espacial.

**Palavras-chave:** *Análise exploratória de dados espaciais, Mortalidade infantil, Municípios.*

---

<sup>1</sup>Universidade Federal de Alfenas - UNIFAL-MG. e-mail: [walefm2@gmail.com](mailto:walefm2@gmail.com)

<sup>2</sup>Universidade Federal de Alfenas - UNIFAL-MG.

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## **O uso do modelo defectivo de Gompertz na modelagem de fração de cura em pacientes portadores de HIV**

Em análise de sobrevivência, modelar o tempo até um evento de interesse ocorrer é de grande importância em estudos de natureza médica e confiabilidade industrial. A literatura apresenta muitas técnicas paramétricas ou não-paramétricas para realizar a modelagem dos tempos de sobrevivência, porém tais técnicas tornam-se limitadas na presença de uma fração de curas na população de estudo, isto é, a presença de pacientes curados ou sobreviventes de longo prazo. Nessa situação a literatura apresenta alguns modelos baseados em misturas de distribuições paramétricas ou alguns modelos de não misturas para modelagem de dados de sobrevivência na presença de fração de curas. Neste trabalho, apresentamos como uma alternativa às técnicas tradicionais existentes, uma nova abordagem para modelar a fração de curas usando modelos defectivos sob o enfoque Bayesiano. Neste caso, foi considerada a distribuição Gompertz que se torna defectiva quando seu parâmetro de escala assume valores negativos. Para ilustrar a performance do modelo adotado na presença de fração de cura, foi considerado um banco de dados relacionado a pacientes portadores do vírus HIV. Os sumários a posteriori foram obtidos utilizando os métodos tradicionais MCMC(Monte Carlo em Cadeias de Markov) por meio do software OpenBUGS.

**Palavras-chave:** análise de sobrevivência, distribuição Gompertz, fração de cura, inferência Bayesiana, vírus HIV.

**Autores:** Milene Regina dos Santos, Ricardo Puziol de Oliveira, Jorge Alberto Achcar e Edson Zangiacomi Martinez.

**Perfil da demanda e padrão das concessões do benefício de prestação continuada (BPC)**

**Eraldo Schunk Silva<sup>1</sup>**  
**Mariana Dias de Campos<sup>2</sup>**  
**Giovanna Lyka Kohatsu<sup>2</sup>**  
**Mirian Ueda Yamaguchi<sup>3</sup>**  
**Ely Mitie Massuda<sup>4</sup>**

**Resumo:** Investigou-se o perfil da demanda e o padrão das concessões do BPC para pessoas com deficiência menores de 16 anos. Foram analisados 331 requerimentos coletados junto ao INSS – Agência Maringá (PR). A avaliação médico-pericial era composta por 83 itens dividida em 19 classes de disfunções. Cada item utilizado para avaliar as funções do corpo é medido numa escala variando de 0 (zero) a 4 (quatro), onde: 0= nenhuma alteração (0 a 4%), 1- alteração leve (5 a 24%), 2= alteração moderada (25 a 49%), 3= alteração grave (50 a 95%) e 4= alteração completa (96 a 100%). O total itens utilizados na avaliação foi resumido em seis classes de disfunções por meio de uma análise de cluster (AC), utilizando-se distância Euclidiana e o método de Ward. Utilizou-se como *input* na AC a matriz de escores fatoriais obtidos por meio da análise fatorial. De 27 BPC concedidos no período, 58,01% foram outorgados em 2016. A grande maioria das demandas estão relacionadas às causas congênitas ou devido às deficiências causadas por doenças. Para a concessão de benefícios devido às disfunções relativas a *Cognição/comunicação* exige-se maior comprometimento das funções do corpo para o grupo etário acima de 7 anos. Para os dois grupos etários, a grande maioria das concessões do BPC, para as disfunções relativas às *Funções Sensoriais*, exigem baixo percentual de itens assinalados com disfunção moderada, grave ou completa, todavia há um percentual maior de concessões para o grupo etário acima de 7 anos.

**Palavras-chave:** *Análise Fatorial, Análise de Cluster, Concessões do Benefício de Prestação Continuada, Disfunções congênitas.*

---

<sup>1</sup>DES – UEM. e-mail: [eraldoschunk@gmail.com](mailto:eraldoschunk@gmail.com)

<sup>2</sup>UNICESUMAR. e-mail: [marianamdccampos@hotmail.com](mailto:marianamdccampos@hotmail.com)

<sup>3</sup>UNICESUMAR/ICETI/PPGPS

<sup>4</sup>UNICESUMAR/ICETI/PPGPS/PPGGCO

# Public spaces as support for physical activity: A study with people over 40 years in the city of Cambé - Paraná

Mariana Ragassi Urbano <sup>1</sup>

Raissa Galvão Bessa <sup>2</sup>

Milena Kanashiro <sup>2</sup>

Ana Carolina Bertin de Almeida Lopes <sup>3</sup>

Ana Julia Righetto <sup>4</sup>

This research aims to study the impact of the public spaces available for the practice of physical activity in the health of the population over 40 years old. The data were collected in the city of Cambé-Paraná-Brazil. The city was divided in six regions: Center, South, Northeast, East, North and West. In all the regions were installed outdoor gyms, that could be used by anyone. Some of the results were that people with less years of education and low income became more satisfied with these public gyms compared to people with more years of education and upper income. The public gyms were installed according the division in censitary regions. Based on the data we propose a new criteria for the instalation of these gyms, ir order to make them available for the people that will make more use of them. All the analysis were performed in software R.

**Keywords:** *Public spaces, Physical activity, Software R, Statistical analysis, Urbanism.*

---

<sup>1</sup>DSTA - UEL. e-mail: [mrurbanouel@gmail.com](mailto:mrurbanouel@gmail.com)

<sup>2</sup>CTU - UEL.

<sup>3</sup>CCS - UEL.

<sup>4</sup>IAPAR.



## Relacionamento da Mortalidade Infantil com a Fecundidade no Semiárido Brasileiro

Mateus Passador Bittencourt de Sá<sup>1 2</sup>

Neir Antunes Paes<sup>2 2</sup>

O principal objetivo consistiu em analisar a relação entre a Taxa de Mortalidade Infantil (TMI) e a Taxa de Fecundidade Total (TFT) para as microrregiões do semiárido brasileiro em 2010. Para tanto, foi utilizada a técnica de regressão clusterwise para as 137 microrregiões com o propósito de agrupar elementos de dados baseando-se na relação linear entre as TMI e as TFT. Para cada cluster encontrado foi feita a validação dos modelos de regressão verificando o cumprimento dos pressupostos através das seguintes técnicas: análise dos resíduos dos modelos, diagnóstico da homocedasticidade pelo teste Breusch-Pagan e o teste de normalidade de Shapiro-Wilks. Fez-se o cálculo das médias ponderadas pelo tamanho da população das microrregiões das TMI estimados pelo Projeto de Busca Ativa do Ministério da Saúde com as estimadas pelo Atlas de Desenvolvimento Humano (ATLAS). As TFT foram extraídas do ATLAS. O modelo de regressão clusterwise apontou três clusters, cujos supostos foram atendidos. A formação dos clusters permitiu a identificação de blocos regionais dentro do semiárido com características distintivas como TMI e TFT mais elevadas em um bloco e variações destas taxas nos demais. Estas características podem auxiliar no monitoramento, otimização e estabelecimento de prioridades inter-regionais quanto aos investimentos na saúde reprodutiva da mulher e da criança, bem como no desenvolvimento socioeconômico, já que este último tem sido reiteradamente associado com estas taxas na literatura.

**Palavras-chave:** *Indicadores demográficos, Regressão clusterwise, Semiárido.*

---

1 DEST – UFPB. e-mail mateuspbsa@gmail.com

2 DEST – Programa de Pós-Graduação em Modelos de Decisão e Saúde, UFPB.

3 Agradecimento ao CNPq pelo apoio na concessão de bolsa de estudo.

## **Insuficiência Renal: Identificando fatores associados utilizando GAMLSS**

**Thiago Gentil Ramires<sup>1</sup>**

**Luiz Ricardo Nakamura<sup>2</sup>**

**Ana Julia Righetto<sup>3</sup>**

**Edwin Moises Marcos Ortega<sup>4</sup>**

**Gauss Moutinho Cordeiro<sup>5</sup>**

Insuficiência renal é um sério problema de saúde pública, em que estudos regionais são necessários para identificar fatores que estão associados com o tempo de vida esperado dos pacientes. Nessa pesquisa, o tempo de vida dos pacientes da região metropolitana de Maringá, Brasil, foram analisados utilizando uma parametrização diferente da usual da distribuição Weibull, na qual contém um parâmetro de locação que é exatamente a média do modelo. Os dados foram coletados no Instituto do Rim de Maringá, contento os pacientes classificados com insuficiência renal crônica, inscritos no programa de hemodiálise desde 1978 até 2010. Devido à falta de hospitais especializados no tratamento dessa doença na região de Maringá, o instituto é o responsável pelo atendimento da maioria dos pacientes com insuficiência renal da grande região metropolitana de Maringá, a qual é formada por 26 cidades, contendo no total aproximadamente 764.000 habitantes. Usualmente a classe de modelos de locação é utilizada para modelar o tempo de vida de pacientes, porém, a desvantagem dessa classe de modelos é que a variabilidade das informações não é modelada utilizando variáveis explanatórias, afetando então a eficiência das estimativas dos parâmetros. Como alternativa, os modelos aditivos generalizados para locação, escala e forma (GAMLSS) permitem que todos os parâmetros do modelo possam ser modelados utilizando variáveis explanatórias. Nós podemos concluir que: quando mais cedo o aparecimento do problema renal, gênero masculino, transplante de rim e positivos anticorpos para hepatite B e C são fatores negativos associados aos tempos de vida dos pacientes. Desta forma, pacientes com tais características requerem atenção especial para que a expectativa de vida possa ser prolongada.

**Palavras-chave:** *GAMLSS, Fator de Risco, Falha Renal.*

---

<sup>1</sup> DAMAT – UTFPR. e-mail [thiagogentil@gmail.com](mailto:thiagogentil@gmail.com)

<sup>2</sup> INE- UFSC

<sup>3</sup> IAPAR

<sup>4</sup> LCE - USP/ESALQ

<sup>5</sup> DE-UFPE

# Spatial Extreme Learning Machines: An Application on Prediction of Disease Counts

Marcos Oliveira Prates <sup>1 3</sup>

<sup>3</sup>

Extreme Learning Machines (ELM) have gained a lot of attention by the Machine Learning community because of its interesting properties and computational advantages. With the increase collection of information nowadays, many sources of data have missing information making statistical analysis harder or unfeasible. In this paper, we present a new model, coined Spatial Extreme Learning Machine (SPELM), that combine spatial modeling with ELM keeping the nice properties of both methodologies and making it very flexible and robust. As explained throughout the text, the SPELM have many advantages in comparison with the traditional ELM. By a simulation study and a real data analysis we present how the SPELM can be used to improve imputation of missing data and uncertainty prediction estimation.

**Palavras-chave:** *Bayesian method, ELM, INLA, Missing Data, Spatial modeling.*

---

<sup>1</sup>DEST - UFMG. e-mail: [marcosop@est.ufmg.br](mailto:marcosop@est.ufmg.br)

<sup>3</sup>Agradecimento ao CNPq, CAPES e FAPEMIG pelo apoio financeiro.

## Uso de um modelo aditivo generalizado para o risco de gravidez na adolescência em Minas Gerais

**Daiane Leite da Roza<sup>1 5</sup>**

**Edson Zangiacomi Martinez<sup>2</sup>**

**Carla Maria Teixeira de Oliveira<sup>3</sup>**

**Maria de Fátima Rodrigues Pereira de Pina<sup>4</sup>**

**Denisa Maria de Melo Vasques de Mendonça<sup>3</sup>**

*Objetivo:* Investigar por meio de uma análise espaço-temporal a associação entre os percentuais de nascidos vivos de mães adolescente (NVMA) e o índice de desenvolvimento humano (IDH), incluindo as três componentes: renda, educação e longevidade.

*Métodos:* Os percentuais de NVMA foram obtidos do sistema de informações de nascidos vivos para o estado de Minas Gerais no período de 2000-2015 e os dados do IDH e suas componentes foram obtidos do programa de desenvolvimento das nações unidas. Um modelo aditivo generalizado (GAM - *Generalized Additive Model*) foi usado para estimar o risco relativo dos percentuais de NVMA em relação ao IDH. Os GAM são uma extensão dos modelos lineares generalizados, eles permitem uma flexibilidade aos efeitos não lineares das covariáveis a serem estimadas na variável resposta dos dados. O modelo assumiu que a distribuição dos percentuais de NVMA segue uma distribuição binomial negativa com função de ligação logarítmica. Para a identificação de agrupamentos espaciais da distribuição geográfica dos percentuais de NVMA foram utilizados o índice global e local de Moran.

*Resultados:* Existe uma associação entre o IDH e suas componentes com os percentuais de NVMA. Os altos valores de risco relativo estão concentrados espacialmente na parte norte do estado de Minas Gerais. Os gráficos indicaram uma relação não linear entre os percentuais de NVMA ao longo dos anos.

*Conclusões:* Existe uma forte dependência espacial de NVMA em Minas Gerais, o que sugere que a localização geográfica exerce um papel fundamental para o seu entendimento.

**Palavras-chave:** *Espaço-Temporal, Modelos Aditivos Generalizados, Risco Relativo.*

---

<sup>1</sup> FMRP – USP. e-mail: daianeroza@usp.br

<sup>2</sup> FMRP – USP.

<sup>3</sup> UNIVERSIDADE DO PORTO.

<sup>4</sup> FIOCRUZ

<sup>5</sup> Agradecimento à CAPES pelo apoio financeiro.

## Esportes

## Estimação de Ratings no Futebol : Campeonato Brasileiro de 2017

Luciano Ribeiro Galvão <sup>1 3</sup>

Júlio Sílvio Sousa Bueno Filho <sup>2 3</sup>

Paulo Henrique Sales Guimarães <sup>2 3</sup>

Dentre os esportes mais praticados no mundo o futebol tem lugar de destaque. Uma das formas de analisar resultados de futebol é considerar os números de gols pró e contra o mandante em cada jogo como se seguissem a distribuição de Poisson. Neste trabalho utilizamos o primeiro turno do Campeonato Brasileiro de Futebol de 2017, por ser esta uma das mais importantes competições do esporte. Ajustamos um modelo linear generalizado aos resultados do primeiro turno e organizamos sistemas de predição para os resultados do segundo turno usando o modelo ajustado com parâmetros de ataque, defesa e mando de campo. As propriedades do ajuste foram boas, mas as da predição variaram muito de acordo com a forma de predição estabelecida.

**Palavras-chave:** *Distribuição de Poisson. Força de Ataque. Força de Defesa. Mando de Campo. Modelo Linear Generalizado.*

---

<sup>1</sup>DES - UFLA. e-mail: [lrgalvao@posgrad.ufla.br](mailto:lrgalvao@posgrad.ufla.br)

<sup>2</sup>DES - UFLA.

<sup>3</sup>Os autores agradecem à FAPEMIG, à CAPES e ao CNPq pelo apoio financeiro.

# Refinamentos de testes de hipóteses em modelos lineares generalizados: uma aplicação a dados esportivos

Tiago Moreira Vargas <sup>1</sup>

Discutimos aqui os principais refinamentos assintóticos de testes de hipóteses aplicados aos modelos lineares generalizados (MLG). Fazemos um estudo detalhado das correções de Bartlett e tipo-Bartlett para os testes baseados nas estatísticas de teste que convergem para a distribuição qui-quadrado sob a hipótese nula em MLG, a saber: Teste da razão de verossimilhanças (Bartlett, 1953) , teste escore (Cordeiro e Ferrari, 1991) e teste gradiente (Vargas, Ferrari e Lemonte, 2013). Ajustamos um modelo de regressão gama a dados de decatlo dos jogos olímpicos de 1988, e comparamos os resultados dos testes baseados nas estatísticas da razão de verossimilhanças, escore, Wald e gradiente não corrigidos com os das versões corrigidas pelo fatores de correção dos testes da razão de verossimilhanças, escore e gradiente. Conclui-se assim, que há mudança de decisão no caso da utilização dos testes corrigidos.

**Palavras-chave:** *Bartlett, tipo-Bartlett, modelos lineares generalizados.*

---

<sup>1</sup>IME - UFG. e-mail: [tiagomvargas@gmail.com](mailto:tiagomvargas@gmail.com)

## Farmácia e Bioquímica



# Análise do Banco de Dados AERS para Explorar Eventos Adversos dos Medicamentos Procinéticos

Jessica Suzana Barragan Alves <sup>1 3</sup>

Andrea Diniz <sup>2 3</sup>

Josmar Mazucheli <sup>2 3</sup>

Naiara C. Aparecido dos Santos <sup>2 3</sup>

Patricia Stülp <sup>2 3</sup>

A classe de medicamentos dos prócinéticos são prescritos em geral para o tratamento de distúrbios da motilidade gastrointestinal, refluxo gastroesofágico. Como o uso dessa classe de medicamentos é muito frequente, encontrar associações entre o aumento do risco de aparecimento de eventos adversos em pacientes debilitados ou susceptíveis é de extrema importância para o aprimoramento de ações de saúde pública. Neste estudo, foi realizada uma análise sobre os eventos adversos dos medicamentos procinéticos a partir o banco de dados do AERS (Adverse Event Report System) no período que corresponde de janeiro à março de 2017.

O (AERS) é uma base de dados de informação aberta criada para apoiar o programa de monitorização pós-comercialização da FDA (*Food and Drug Administration*) dos Estados Unidos, para todos os medicamentos e produtos terapêuticos aprovados. Seu principal objetivo é coletar e armazenar relatórios de segurança para o monitoramento da saúde da comunidade (YILDIRIM, 2015).

Para a pesquisa foram usados os seguintes medicamentos que compõe a classe dos procinéticos: bomoprida, cisaprida, metoclopramida, domperidona, meclizina, ondansetrona, alosetrona e tegaseroda. O objetivo deste trabalho foi explorar associações desconhecidas entre os eventos adversos a idade, o sexo do paciente e os desfechos dos eventos adversos. Para isso foi utilizado regras de associação e o algoritmo a priori.

A pesquisa resultou em 591 relatórios onde, verificou-se o evento adverso e o desfecho mais frequentes, foi construído regras de associações para responder o objetivo proposto.

**Palavras-chave:** FAERS, Farmacovigilância, Procinéticos, Regra de Associação.

---

<sup>1</sup>PBE - UEM. e-mail: [jessicasbarragan@hotmail.com](mailto:jessicasbarragan@hotmail.com)

<sup>2</sup>PBE - UEM.

<sup>3</sup>Agradecimento ao CNPQ pelo apoio financeiro.

## Genética e Melhoramento

## AMMI e GGE Biplot para interação Genótipo × Ambiente: uma análise de clusters baseada em medóides

Anderson Cristiano Neisse<sup>1</sup>

Jhessica Letícia Kirch<sup>2</sup>

Kuang Hongyu<sup>3</sup>

A presença de interação genótipo-ambiente (GEI) influencia a produção tornando a seleção de cultivares superiores um processo complexo. Ambos modelos AMMI (additive main-effects and multiplicative interaction) e GGE Biplot (genotype main-effects and genotype-environment interaction) são amplamente utilizados para análise de dados de ensaios multi-ambientes (Dados MET). O AMMI considera somente o efeito GEI na análise, já o GGE Biplot considera também efeitos principais de genótipos de forma aditiva. Ambos os modelos são baseados em biplots e consequentemente interpretações baseadas em biplots não escalam muito bem além de duas variáveis, o que acontece sempre que os dois primeiros componentes principais não capturam variância suficiente dos dados. Este estudo propõe uma abordagem para estes casos, baseada na combinação de análise de cluster e medóides para seleção de representantes a serem utilizados na análise. Também interpreta os resultados de ambos os modelos AMMI e GGE de forma a comparar decisões obtidas de ambos. Os dados foram fornecidos pelo International Maize and Wheat Improvement Center (CIMMYT) provenientes do 14th Semi-Arid Wheat Yield Trial (SAWYT), experimento com 50 genótipos de trigo (*Triticum aestivum*) adaptado a baixos níveis de chuva. O experimento foi realizado em 36 ambientes dentre 14 países. Por meio da análise de clusters combinada com medóides foram escolhidos 25 genótipos e 6 ambientes. Ambos os modelos mostraram equivalência para a caracterização dos dados, aumentando a confiabilidade na seleção de cultivares superiores e ambientes de teste. Após análise conjunta os genótipos G8 e G45 foram avaliados como os mais próximos de um genótipo ideal.

**Palavras-chave:** *Interação genótipo × ambiente; adaptabilidade e estabilidade; modelo AMMI; GGE Biplot; ensaios multi-ambientes; análise de cluster; medóides.*

---

<sup>1</sup> PPESTBIO – UFV. e-mail [a.neisse@gmail.com](mailto:a.neisse@gmail.com)

<sup>2</sup> LCE – ESALQ/USP.

<sup>3</sup> DEST – ICET - UFMT.

## **Análise multiambientes de genótipos de milho usando modelo AMMI-Bayesiano**

**Luciano Antonio de Oliveira**<sup>1 2 3</sup>

**Cristian Tiago Erazo Mendes**<sup>1 3</sup>

**Alessandra Querino da Silva**<sup>2 3</sup>

**Carlos Pereira da Silva**<sup>1 3</sup>

**Marcio Balestre**<sup>1 3</sup>

O objetivo deste trabalho foi analisar adaptabilidade e estabilidade de 55 genótipos de milho em 15 ambientes distintos utilizando o Modelo de Efeitos Principais Aditivos e Interação Multiplicativa (AMMI), sob a ótica da inferência bayesiana (AMMI-Bayesiano). O delineamento utilizado foi o de blocos completos casualizados com três repetições e a variável analisada foi produtividade de espigas em  $t\ ha^{-1}$ . Os resultados das análises evidenciaram a grande flexibilidade do AMMI-Bayesiano para a incorporação de inferência aos parâmetros do modelo, incluindo os escores genotípicos e ambientais que descrevem o efeito da interação entre genótipos e ambientes (GE) na representação biplot, relativa aos dois primeiros componentes principais. As regiões de credibilidade bivariadas incorporadas aos escores no biplot permitiram identificar subgrupos homogêneos de genótipos e ambientes, bem como aqueles que não possuem contribuição significativa para interação. Essas regiões também permitiram maior confiabilidade para o estudo da adaptabilidade de genótipos a ambientes específicos, que são de grande interesse em programas de melhoramento de plantas. Dessa forma, o ranqueamento das médias à posteriori, para efeito de genótipos, utilizando regiões de máxima densidade a posteriori (HPD), aliadas às informações obtidas por meio das regiões de credibilidade, para os escores genotípicos e ambientais, possibilitou a identificação dos melhores genótipos, em relação a característica em análise.

**Palavras-chave:** *Adaptabilidade, Análise Bayesiana, Estabilidade, Milho.*

---

<sup>1</sup> DES – UFLA. e-mail: lucianoantonio@ufgd.edu.br

<sup>2</sup> FACET– UFGD

<sup>3</sup> Agradecimento a FAPEMIG pelo apoio financeiro.

# Applications of Weighted AMMI Model

Danilo Augusto Sarti <sup>1</sup>

Carlos Tadeu dos Santos Dias <sup>2</sup>

This paper aims to describe the use of weighted AMMI model to evaluate the stability of 8 genotypes of corn in a block design conducted in a second season corn environment. Genotype x Environment interactions will be explored with biplots. We explore also aspects of yeild versus stability using PCA analysis and Biplots.

**Palavras-chave:** *Multivariate, W-AMMI, Genetics.*

---

<sup>1</sup>LCE - USP/ESALQ. e-mail:danilo.estat@gmail.com

<sup>2</sup>LCE - USP/ESALQ. e-mail:ctsdias@usp.br

## **Aprendizado de máquina na predição genômica da resistência à ferrugem alaranjada em café arábica.**

**Ithalo Coelho de Sousa<sup>1 5</sup>**

**Moysés Nascimento<sup>2 5</sup>**

**Cosme Damião Cruz<sup>3 5</sup>**

**Ana Carolina Nascimento<sup>2 5</sup>**

**Eveline Teixeira Caixeta<sup>4 5</sup>**

A seleção genômica (SG) foi proposta como uma forma de aumentar a eficiência e acelerar o melhoramento genético. Metodologias estatísticas têm sido utilizadas em SG para a predição do mérito genético, como por exemplo o RR-BLUP e BLASSO. Entretanto, tais metodologias exigem algumas pressuposições a respeito dos dados tais como normalidade da distribuição dos valores fenotípicos. Além disto, a presença de fatores complicadores como por exemplo a epistasia e dominância atrapalham a utilização destas. Uma alternativa interessante são os modelos baseados em aprendizagem de máquina, como por exemplo às Redes Neurais (RNA) e Árvores de Decisão (AD) que podem incorporar de maneira natural a epistasia e a dominância no modelo além de não exigirem pressuposições quanto a distribuição dos valores fenotípicos. O objetivo deste trabalho foi utilizar o BLASSO, RNA e AD para predição da resistência a ferrugem no café arábica. Além disso, as metodologias foram utilizadas para identificar a importância dos marcadores relacionados a característica de interesse. Foram utilizados dados da resistência a ferrugem do café de 245 plantas derivadas do cruzamento do Híbrido de Timor e do Catuaí Amarelo, genotipados para 137 marcadores. A RNA obteve maior eficácia, em termos da taxa de erro aparente, seguida do BLASSO e da AD, respectivamente, na predição do caráter em estudo. Relativamente ao custo computacional a AD apresentou um melhor desempenho seguido do BLASSO e da RNA, respectivamente. Ademais, o BLASSO e a AD foram capazes de identificar marcadores importantes para a característica de estudo com base em resultados da literatura.

---

<sup>1</sup> DET – UFV. e-mail [ithalo.coelho@gmail.com](mailto:ithalo.coelho@gmail.com)

<sup>2</sup> DET - UFV

<sup>3</sup> DBG - UFV

<sup>4</sup> EMBRAPA Café

<sup>5</sup> Agradecimento à CAPES e a FAPEMIG pelo apoio financeiro.

**Palavras-chave:** Árvore de decisão, Rede neural, BLASSO.

## Avaliação da capacidade preditiva do modelo AMMI-Bayesiano

Cristian Tiago Erazo Mendes <sup>1 4</sup>

Luciano Antonio de Oliveira <sup>1 2 4</sup>

Carlos Pereira da Silva <sup>1 4</sup>

Andrezza Kellen Alves Pamplona <sup>1 3 4</sup>

Márcio Balestre <sup>1 4</sup>

A análise de dados multi-ambientais (MET) tem assumido grande importância no contexto dos programas de melhoramento de plantas, em que o intuito é permitir um estudo criterioso da interação entre genótipos e ambientes (GEI). Nesse sentido, estudos recentes têm mostrado vantagens sugestivas da utilização do método bayesiano aplicado ao modelo de efeitos aditivos e interação multiplicativa (AMMI-Bayesiano) em relação à abordagem tradicional (AMMI, considerando efeitos como fixos). O presente trabalho teve por objetivo verificar a robustez na capacidade preditiva do modelo AMMI-Bayesiano. Para tanto foi simulado um conjunto de dados com 20 genótipos (G1-G20), avaliados em cinco ambientes (E1-E5), em um delineamento de blocos ao acaso com três repetições. Foram realizados desbalanceamentos aleatórios os níveis de 10%, 33% e 50% de perda de genótipos nos ambientes. Para avaliar a capacidade preditiva, foram utilizadas a estatística PRESS (prediction error sum square) e a correlação entre o valor predito e observado. As análises foram conduzidas com o uso do software R. Os resultados mostraram que o modelo analisado é robusto sob desbalanceamento e, portanto flexível para analisar dados MET superando limitações do modelo AMMI tradicional, em lidar com desbalanceamento nos dados ou perda de parcelas. Desta forma, esse método é promissor e pode facilitar significativamente o trabalho de melhoristas na avaliação de dados MET.

**Palavras-chave:** *dados multi-ambientais, método bayesiano, modelo AMMI.*

---

<sup>1</sup>DES - UFLA. e-mail: [tiagoerazo@gmail.com](mailto:tiagoerazo@gmail.com)

<sup>2</sup>FACET - UFGD.

<sup>3</sup>IFTM.

<sup>4</sup>Agradecimentos a CAPES e FAPEMIG pelo apoio financeiro.



## Avaliação da performance do método RKHS para predição de GEBV usando as informações de pedigree e marcadores

Andrew de Paula Ribeiro <sup>1 3</sup>

Geraldo Magela da Cruz Pereira <sup>2 3</sup>

Sebastião Martins Filho <sup>2</sup>

A predição acurada de valores genéticos genômicos (GEBVs) é um fator importante para o melhoramento genético vegetal e animal. Há tempos, o mérito genético de indivíduos vem sendo predito, pela utilização de informações fenotípicas e de genealogia. Com o advento da seleção genômica ampla, tornou-se possível a utilização de marcadores, dispersos pelo genoma, permitindo selecionar indivíduos superiores de forma acurada. Neste trabalho avaliou-se a performance do método RKHS para a predição de GEBVs, variando a densidade dos marcadores e considerando as fontes de informação de pedigree (P) e marcadores (M). Para isso, foram utilizados dados simulados (QTL-MAS Workshop) de uma população de gado leiteiro, constituída por 4100 animais, genotipados por 10 mil marcadores SNPs, sendo a característica de interesse a produção de leite. Diferentes densidades de marcadores foram obtidas, retirando-se aleatoriamente amostras de tamanho 1,5k, 4k, 4,5k, 6k, 7,5k, 9k e 10k. Percebe-se que os modelos M-RKHS e MP-RKHS, apresentaram uma capacidade preditiva superior à obtida pelo P-RKHS. A superioridade do MP-RKHS, sobre o M-RKHS é nítida para baixas densidades, com o aumento do número de marcadores, os métodos tendem a uma mesma capacidade preditiva. Para 1k, maiores valores de variância residual foram obtidos pelos métodos M-RKHS e P-RKHS, já para a densidade de 6k, nota-se uma queda drástica da variância residual do M-RKHS. Em suma, nota-se que a performance do método RKHS, é influenciada pelo tipo de informação utilizada para a predição e pela densidade dos marcadores, sendo a densidade de 6k suficiente para uma boa predição.

**Palavras-Chave:** *Capacidade preditiva, Método Semiparamétrico, Seleção genômica ampla.*

---

<sup>1</sup>DFT - UFV, e-mail: [andrew.ribeiro@ufv.br](mailto:andrew.ribeiro@ufv.br)

<sup>2</sup>DET-UFV.

<sup>3</sup>Agradecimentos à CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## **Avaliação de Endogamia em Populações Seleccionadas e Não- Seleccionadas Via Simulação Monte Carlo com Auxílio do Software R**

**Diego Duarte Varela<sup>1 3</sup>**

**Crysttian Arantes Paixão<sup>2 3</sup>**

**Heloísa Maria de Oliveira<sup>2 3</sup>**

O melhoramento genético animal se intensificou devido aos novos estudos que utilizam principalmente modelos mistos biométricos. Esses modelos descrevem e simulam rebanhos através das características de raça, sexo, idade, peso, entre outras. Este trabalho tem por objetivo estudar a interação entre os indivíduos e sua evolução dentro da simulação via Monte Carlo com o auxílio do software R, permitindo avaliar os índices de endogamia, distribuição de genes na população e peso por meio do modelo animal aditivo. Os animais foram caracterizados pelo genótipo “A” e “a” em cada alelo para descrever o peso. Ainda como características considerou-se o peso, sexo, idade, parentesco quando as simulações de cruzamento foram realizadas. Os dados gerados condizem com o esperado em populações reais, encontrados na literatura, já que, apesar da ocorrência de acasalamentos entre parentes próximos, os alelos são passados entre as populações seleccionadas entre os machos-base da mesma. Com base nesse estudo, pode-se afirmar que a população não-seleccionada apresentou uma estabilização do peso e aumento gradativo da endogamia, visto que o equilíbrio de Hardy-Weinberg não aconteceu devido à ocorrência de acasalamentos entre parentes por uma distribuição não efetiva dos alelos. Já na população seleccionada, ocorreu aumento progressivo do peso e diminuição seguida de estabilização da endogamia. As simulações relatadas, cujas representatividade e aleatoriedade altas não são encontradas a campo, são de alto teor científico e podem ser utilizadas tais quais ferramentas do melhoramento genético animal.

**Palavras-chave:** *Simulação, Melhoramento Animal, Endogamia.*

---

<sup>1</sup> CCR – UFSC. Graduando do curso de Medicina Veterinária da Universidade Federal de Santa Catarina. E-mail: [diego.duarte.varela@grad.ufsc.br](mailto:diego.duarte.varela@grad.ufsc.br).

<sup>2</sup> CCR – UFSC. Professor Adjunto na Universidade Federal de Santa Catarina.

<sup>3</sup> Agradecimento ao Grupo de Estudos de Estatística Teórica, Aplicada e Computacional pelo apoio.

## **Caracterização dos atributos sensoriais de bebida de Café arábica de acordo com o ambiente**

**Ivan de Paiva Barbosa<sup>1 5</sup>**

**Ithalo Coelho de Sousa<sup>2 5</sup>**

**Renato Domiciano Silva Rosado<sup>2 5</sup>**

**Antônio Carlos Baião de Oliveira<sup>3 5</sup>**

**Cosme Damião Cruz<sup>4 5</sup>**

No mercado de cafés especiais devemos levar em consideração as diferentes regiões produtoras, o nível de diferenciação do café e a qualidade sensorial para agregar valor ao produto final. Assim, objetivou-se caracterizar os atributos sensoriais de genótipos de *Coffea arabica* de acordo com o ambiente na região das Matas de Minas. Os experimentos foram instalados em três municípios na região das Matas em MG. O delineamento foi em blocos casualizados com duas repetições e onze genótipos. As amostras de frutos do tipo cereja foram avaliadas segundo o protocolo da “Specialty Coffee Association of America”. Os dados foram submetidos a ANOVA e a variabilidade relativa dos atributos sensoriais pelo critério proposto por Singh (1981) e uma análise descritiva por meio do perfil sensorial. Os atributos sabor no município de Paula Cândido, fragrância/aroma, corpo e finalização em Senhora de Oliveira e corpo em Araponga, apresentaram baixa variabilidade relativa podendo ser considerados características comuns entre os cafés despulpados produzidos nos respectivos ambientes. Já considerando a análise conjunta dos ambientes observamos que as características de menores variabilidade relativa foram aquelas de mais baixa herdabilidade e de menor relação entre os coeficientes de variação genética e ambiental. Na análise por ambiente, de forma geral, o corpo e a fragrância/aroma, nesta ordem, foram os atributos que apresentaram em média, baixa variabilidade relativa, além de uma tendência para notas altas. Desta forma, o corpo e a fragrância podem ser considerados atributos comum entre os cafés despulpados produzidos na região das Matas de Minas.

**Palavras-chave:** *Qualidade sensorial, Coffea arabia, Cafés especiais.*

<sup>1</sup> DBG – UFV. e-mail ivan.barbosa@ufv.br

<sup>2</sup> DET – UFV.

<sup>3</sup> EPAMIG – MG.

<sup>4</sup> DBG – UFV

<sup>5</sup> Agradecimento ao CNPq e a FAPEMIG pelo apoio financeiro.

**Classificação de genótipos de Arroz Irrigado quanto a Adaptabilidade e Estabilidade em Minas Gerais****Antônio Carlos da Silva Júnior<sup>1 2</sup>****Weverton Gomes da Costa<sup>12</sup>****Ivan de Paiva Barbosa<sup>12</sup>****Plínio César Soares<sup>23</sup>****Cosme Damião Cruz<sup>24</sup>**

Estudos de adaptabilidade e estabilidade são de grande importância para auxiliar a classificação de cultivares uma vez que permitem obter informações detalhadas do comportamento dos genótipos avaliados em uma região. Portanto, o objetivo deste trabalho foi avaliar a adaptabilidade e estabilidade de genótipos de arroz irrigado por meio das metodologias de Eberhart e Russell (1966) e dos centroides múltiplos (2015), para auxiliar na recomendação de cultivares de arroz irrigado no Estado de Minas Gerais. Foram avaliados, quanto à produtividade de grãos ( $\text{Kg.ha}^{-1}$ ), 18 genótipos de arroz irrigado incluindo cinco testemunhas. Os ensaios foram conduzidos nos municípios de Lambari, Leopoldina e Nova Porteirinha, nos anos agrícolas de 2012/2013, 2013/2014, 2014/2015 e 2015/2016, perfazendo 12 ambientes. Foram realizadas análises de variância conjunta e de adaptabilidade e estabilidade segundo métodos Eberhart e Russell (1966) e dos centroides múltiplos (2015). Observou-se que a utilização destas metodologias é uma alternativa viável para classificação de genótipos. Além disso, os genótipos comportaram-se de forma diferente quanto à estabilidade e adaptabilidade nos diferentes ambientes. Ambas as metodologias, identificaram as linhagens BRA 02691 e MGI 0607-1 como promissoras para serem lançadas como cultivares, entretanto, o método do centroide demonstrou maior sensibilidade em relação ao método de Eberhart e Russell.

**Palavras-chave:** *Oryza sativa* L., Biometria, Centroide Múltiplo, Eberhart e Russel.

<sup>1</sup> Departamento de Biologia Geral – UFV. e-mail [antonio.c.junior@ufv.br](mailto:antonio.c.junior@ufv.br)

<sup>2</sup> Agradecimento ao CNPq, FAPEMIG e Capes pelo apoio financeiro.

<sup>3</sup> Pesquisador – EPAMIG/Viçosa-MG/Brasil. Bolsista BIPTD FAPEMIG

<sup>4</sup> Professor Titular- Departamento de Biologia Geral- UFV/Viçosa-MG/Brasil. Bolsista CNPq

## Comparação de métodos bayesianos para a seleção genômica de características ordinais em dados simulados

Geraldo Magela da Cruz Pereira<sup>1 3</sup>

Andrew de Paula Ribeiro<sup>2 3</sup>

Sebastião Martins Filho<sup>2</sup>

A acurácia da predição genômica para características contínuas e normalmente distribuídas, depende de diversos fatores, dentre eles: número de marcadores, modelo e herdabilidade da característica. Para características de natureza ordinal, esta dependência deve ser melhor investigada. Neste trabalho objetivou-se comparar os métodos bayesianos BayesC e BLASSO, para a predição genômica de características fenotípicas ordinais, considerando 2 densidades de marcadores (3k e 6k), dois valores de herdabilidade (0,56 e 0,66) e a categorização do fenótipo em 2, 3 ou 4 classes. Para avaliação dos métodos foram utilizadas informações genotípicas e fenotípicas, de uma população de 3000 animais, genotipados por 6k marcadores (QTL-MAS Workshop). A capacidade preditiva (CP) e o viés dos modelos foram acessados por validação cruzada (5-fold), sendo a CP dada pela correlação de Pearson entre o valor genético genômico predito e o verdadeiro valor genético genômico. Uma ANOVA foi realizada para avaliar qual fator influencia significativamente a CP. Os resultados da análise de variância mostram que o número de marcadores, herdabilidade e o tipo categorização da resposta, afetam significativamente ( $p < 0,05$ ) a CP. A herdabilidade apresentou o maior valor F, o que implica em uma maior influência sobre a CP. A escolha do modelo foi pouco influente na predição de valores. As metodologias mostraram-se altamente viesadas para a herdabilidade de 0,56, uma média de 3,22. Para herdabilidade de 0,66, o viés médio foi de 1,45. Por fim, nota-se que a acurácia da predição genômica é fortemente dependente da herdabilidade, da densidade de marcadores e do número de classes.

**Palavras-chave:** *Capacidade preditiva, Resposta ordinal, Seleção genômica ampla.*

---

<sup>1</sup> DET – UFV, e-mail: [geraldomcpereira@gmail.com](mailto:geraldomcpereira@gmail.com)

<sup>2</sup> DET – UFV.

<sup>3</sup> Agradecimentos à CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## **Crítérios de decisão do número ótimo de componentes independentes aplicados à modelos aditivo-dominante**

**Jaquicele Aparecida da Costa<sup>1 5</sup>**

**Camila Ferreira Azevedo<sup>2 5</sup>**

**Marcos Deon Vilela de Resende<sup>3 5</sup>**

**Fabyano Fonseca e Silva<sup>4 5</sup>**

**Moysés Nascimento<sup>2 5</sup>**

A seleção genômica se baseia na estimação dos efeitos de uma alta densidade de marcadores no fenótipo de interesse. No entanto, a análise estatística dos dados genômicos apresenta desafios relacionados à alta dimensionalidade e a multicolinearidade entre estes marcadores, e por este motivo, é fundamental a busca de metodologias estatísticas que os contemplem. Dentre elas, destacam-se a Regressão via Componentes Principais (PCR) e Regressão via Componentes Independentes (ICR). Estas metodologias exigem a determinação do número ótimo de componentes (combinações lineares das variáveis originais) que serão incluídos no modelo. Existem proposições de critérios de escolha para a PCR e para a ICR há somente o critério exaustivo (escolhe o número de componentes associado a uma maior acurácia), entretanto, demanda elevado esforço computacional. Além disso, estudos recentes revelam a importância dos modelos aditivo-dominante, que possibilitam avaliações genéticas relacionadas ao valor genômico aditivo, devido à dominância e total. Diante do exposto, este trabalho avaliou, em dados simulados, três critérios de decisão no qual o número ótimo de componentes independentes equivale ao número de componentes que conduz a: PCR a maior acurácia, PCR a menor viés e 80% da variação total explicada pelos componentes principais. As populações simuladas, constituídas por 1000 indivíduos e 2000 marcadores, contemplaram dois níveis de herdabilidade e duas arquiteturas genéticas. Verificou-se que o critério mais acurado determina que o número de componentes independentes equivale ao número de componentes principais que conduz a PCR a uma maior acurácia associada ao valor genômico devido à dominância.

**Palavras-chave:** *componentes independentes, métodos de redução dimensional, modelos aditivo-dominante.*

<sup>1</sup> Programa de Pós-graduação em Estatística Aplicada e Biometria – UFV. e-mail [jaquicele.costa@ufv.br](mailto:jaquicele.costa@ufv.br)

<sup>2</sup> Departamento de Estatística – UFV. e-mail [camila.azevedo@ufv.br](mailto:camila.azevedo@ufv.br), [moysesnasm@ufv.br](mailto:moysesnasm@ufv.br)

<sup>3</sup> Embrapa Floretas/Departamento de Engenharia Florestal. e-mail [marcos.deon@gmail.com](mailto:marcos.deon@gmail.com)

<sup>4</sup> Departamento de Zootecnia – UFV. e-mail [fabyanofonseca@ufv.br](mailto:fabyanofonseca@ufv.br)

<sup>5</sup> Agradecimentos a Fundação Arthur Bernardes (FUNARBE), CAPES e FAPEMIG, pelo apoio financeiro.

## Importância da população de treinamento das redes neurais artificiais para solução de problemas classificatórios

Iara Gonçalves dos Santos<sup>1 3</sup>

Cosme Damião Cruz<sup>1 3</sup>

Weverton Gomes da Costa<sup>1 3</sup>

Alexandre Gomes Ferraz<sup>1 3</sup>

Reinaldo de Paula Ferreira<sup>2 3</sup>

O objetivo desse trabalho foi avaliar eficiência das redes neurais (RNAs) treinadas a partir de dados ampliados de poucos cortes de alfafa para classificação de um novo conjunto de dados. Setenta e sete genótipos de alfafa foram previamente classificados em ótimos, bons, regulares e ruins segundo os escores do índice de Taí para os caracteres de produção e valor nutritivo. Foram estabelecidas diferentes topologias de redes para cada grupo de caracteres, que foram treinadas com dados ampliados de três cortes e posteriormente, utilizadas para classificar dados reais de um novo corte. Considerando a classificação dos genótipos quanto aos caracteres de produção, observou-se altas taxas de erro aparente tanto no treinamento quanto na validação. Para o grupo de caracteres de valor nutritivo foram observadas pequenas taxas de erro tanto no treinamento quanto na validação. Entretanto as baixas taxas de erro não se mantiveram na utilização com um novo conjunto de dados, indicando a perda da habilidade de generalização, possivelmente pelo *overfitting* que pode ter ocorrido. O baixo desempenho das redes pode ser explicado uma vez que os limites mínimos das variáveis no conjunto de treinamento estão bem aquém daqueles encontrados no conjunto de teste. Dessa forma, conclui-se que a quantidade e diversidade dos dados do conjunto de treinamento é importante para manter a capacidade de generalização das RNAs.

**Palavras-chave:** RNAs, inteligência artificial, aprendizado supervisionado.

---

<sup>1</sup> DBG – UFV. e-mail iara.santos@ufv.br

<sup>2</sup> Embrapa Pecuária Sudeste

<sup>3</sup> Agradecimento ao CNPq e a Fapemig pelo suporte financeiro.

## Modelo funcional bayesiano para seleção genômica: métodos RR-BLUP e Bayes A

Andrezza Kellen Alves Pamplona <sup>1 4</sup>

Ernandes Guedes Moura <sup>2 4</sup>

Marcio Balestre <sup>3 4</sup>

Júlio Sílvio de Sousa Bueno Filho <sup>3 4</sup>

Uma técnica de redução de dimensionalidade e multicolinearidade foi proposta por Hu et (2012) por meio do modelo denominado genoma contínuo, no qual se divide o genoma em *bins* e assume-se que a expressão gênica de um *locus* é função desconhecida da posição no genoma. O interesse é estimar tal função e usá-la para prever o valor genético genômico de novos indivíduos. Assim, os autores utilizaram a média dos *bins* como medida de informação. Uma alternativa foi proposta por Moura (2017), na qual se atribuem pesos para os efeitos dentro de um *bin* através da frequência relativa com que cada marcador é assumido possuir efeito dentro de um processo estocástico MCCM (Monte Carlo Cadeias de Markov). Com isso, este trabalho tem por objetivo apresentar a adaptação dos métodos RR-BLUP e Bayes A sob o modelo funcional bayesiano em *bins* e compará-los com os métodos originais quanto à capacidade preditiva. Um genoma com 12150 marcadores foi simulado e os métodos foram analisados em dois cenários (oligogênico e poligênico) com duas configurações de *bins* (10 e 150 *bins*) para as herdabilidades 0,2 e 0,8. Para o cenário oligogênico, os métodos adaptados RR-BLUP e Bayes A utilizando 10 *bins* foram mais acurados do que os respectivos métodos originais. Já para o cenário poligênico, RR-BLUP e Bayes A originais foram mais acurados que os respectivos métodos com *bins*. O modelo funcional bayesiano mostrou-se atrativo na identificação de regiões causais e alta flexibilidade de análise, podendo ser adaptável a qualquer método usual de seleção genômica.

**Palavras-chave:** *Redução de dimensionalidade, Modelo genoma contínuo, Valor genômico, Genoma em bins.*

---

<sup>1</sup>IFTM - campus Uberaba. Doutoranda DES - UFLA. e-mail: [andrezzapamplona@iftm.edu.br](mailto:andrezzapamplona@iftm.edu.br)

<sup>2</sup>IFMA - campus São João dos Patos. Doutorando DES - UFLA.

<sup>3</sup>DES - UFLA.

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.



## Modelo GGE ponderado no estudo da interação entre QTL e ambiente

Tatiana Oliveira Gonçalves de Assis <sup>1</sup>

Carlos Tadeu dos Santos Dias <sup>2</sup>

Paulo Canas Rodrigues <sup>3 4</sup>

Visto que características genéticas das culturas agrícolas podem ser afetadas pelo ambiente, interferindo na produtividade, é importante a pesquisa buscar o aprimoramento das ferramentas capazes de analisar a interação entre genótipo e ambiente (IGA) e interação entre locus de característica quantitativa (QTL) e ambiente (IQA) do cultivar. Neste trabalho, o modelo GGE (Genotype main effects + Genotype environment interaction) será usado na detecção de QTL e no estudo da interação entre QTL e ambiente. Será também proposto um modelo W-GGE Biplot (GGE-Biplot ponderado) com um algoritmo de pesos. Sendo o modelo W-GGE uma generalização do modelo GGE capaz de levar em conta a heterogeneidade de variância do erro, uma vez que se houver uma forte IGA nos dados, é de se esperar ter variâncias de erro heterogêneas, e isso não é levado em conta pelo modelo GGE padrão. Por fim, uma comparação com os resultados obtidos com o modelo AMMI (Additive Main effects and Multiplicative Interaction) e W-AMMI (Rodrigues et al., 2014) será feita com os modelos GGE e W-GGE-Biplot.

**Palavras-chave:** *Modelo GGE, Modelo AMMI, Interação entre genótipo e ambiente, detecção de QTLs, Interação entre QTL e ambiente.*

---

<sup>1</sup>ESALQ - USP. e-mail: [tatiassis@usp.br](mailto:tatiassis@usp.br)

<sup>2</sup>ESALQ - USP.

<sup>3</sup>DEst - UFBA.

<sup>4</sup>University of Tampere, Finland.

## Modelos para predição de área foliar de espécies forrageiras

Hygor Amaral Santana<sup>1 2</sup>

Brunna Rithielly Rezende<sup>1 2</sup>

Wilhan Valasco dos Santos<sup>1 2</sup>

Anderson Rodrigo da Silva<sup>1 2</sup>

A área foliar é uma variável importante em estudos fitofisiológicos, por estar relacionada com a taxa fotossintética e outros processos. Assim, surge a necessidade de métodos rápidos e precisos para estimar a área foliar. O objetivo do trabalho foi ajustar modelos de regressão para predizer a área foliar de seis espécies de leguminosas forrageiras, individualmente, com base em imagens digitais analisadas com o pacote LeafArea, software R. No experimento, avaliou-se 100 folhas aleatoriamente retiradas das seguintes espécies: *C. juncea*, *C. ensiformis*, *C. cajan*, *D. lablab*, *M. cinereum* e *M. aterrima*, sendo utilizadas 70 folhas para ajuste dos modelos e o restante (30) para validação do modelo, tendo como variáveis explicativas o comprimento e a largura do folíolo central. Imagens digitais de cada folha foram processadas para estimar a área foliar. Essas estimativas foram utilizadas para ajustar os modelos, linear e não linear, de predição da área foliar. Para as seis espécies, o modelo linear múltiplo de segundo grau completo pode ser usado para predizer a área foliar em função do comprimento e da largura do folíolo central, apresentando  $R^2 = 0,98$  e erro médio absoluto inferior a 9%. Nos modelos, o efeito da largura do folíolo é, em geral, maior que o comprimento. O pacote R LeafArea mostrou ser uma ferramenta muito eficiente na estimativa da área foliar através da execução do software ImageJ®, com alta precisão e fácil calibração.

**Palavras-chave:** *Crotalária*, *Fabaceae*, *Imagem digital*.

<sup>1</sup> Instituto Federal Goiano – Campus Urutaí. E-mail: hygor\_amaral@hotmail.com

<sup>2</sup> Agradecimento ao IF Goiano pelo apoio financeiro.

## Modelos mistos para experimentos provenientes ao melhoramento de plantas

Alessandra dos Santos <sup>1 2</sup>

Clarice Garcia Borges Demétrio <sup>3</sup>

Renata Alcarde Sermarini <sup>3</sup>

Chris Brien <sup>4</sup>

Um dos principais objetivos dos programas de melhoramento de plantas é selecionar, ao final de estudos, as melhores linhagens testadas. Elas devem ser mais produtivas que as variedades já encontradas no mercado. Para tal, é comum se iniciar o processo de seleção com mais de 100 novos indivíduos em áreas experimentais consideradas grandes devido ao número de parcelas necessárias. Assim, nem sempre o modelo tradicional de análise de variância representa corretamente os efeitos ambientais e de tratamentos (linhagens) existentes. A competição entre parcelas de diferentes linhagens é um dos efeitos que pode acarretar sério viés na análise. Além disso, parcelas mais próximas tendem a ter um efeito ambiental mais parecido que parcelas mais distantes. Por isso, assumir erros independentes e identicamente distribuídos também pode proporcionar resultados não realísticos. Nesse trabalho são apresentados modelos mistos que levam em consideração diferentes estruturas de covariância no efeito ambiental e de tratamentos. Discussões são feitas com base em um experimento em blocos casualizados com 220 novas linhagens de cana-de-açúcar do programa de melhoramento do Instituto Agrônômico de Campinas.

**Palavras-chave:** *Processo autoregressivo, Efeito de competição, Seleção genética.*

---

<sup>1</sup>ESALQ/USP. e-mail: [alessandra.s@usp.br](mailto:alessandra.s@usp.br)

<sup>2</sup>Agradecimento ao CNPq pelo apoio financeiro.

<sup>3</sup>ESALQ/USP.

<sup>4</sup>UNISA/Australia.

## Predição genômica combinando Modelos Mistos e Modelos Bayesianos

Leísa Pires Lima<sup>1 6</sup>

Camila Ferreira Azevedo<sup>2 6</sup>

Marcos Deon Vilela de Resende<sup>3 6</sup>

Fabyano Fonseca e Silva<sup>4 6</sup>

José Marcelo Soriano Viana<sup>5 6</sup>

A Seleção Genômica Ampla (GWS) faz uso de informações diretas do DNA por meio dos marcadores moleculares e visa estimar o efeito de cada marcador no fenótipo. Seu sucesso se dá pela escolha de modelos genéticos-estatísticos confiáveis. O método G-BLUP (*Genomic Best Linear Unbiased Predictor*) é tradicionalmente aplicado a GWS e o BLASSO (*Bayesian Least Absolute Shrinkage and Selection Operator*) vêm se mostrando mais eficiente na predição genômica. Desta forma, o método proposto, denominado Bayes Híbrido, consiste em combinar os valores genômicos estimados via G-BLUP e via BLASSO. Neste método, a informação advinda do G-BLUP é considerada na distribuição *a posteriori* do BLASSO por meio da distribuição *a priori* dos componentes de variância (por meio dos hiperparâmetros). Desta forma, espera-se que o BLASSO utilizando uma distribuição *a priori* altamente informativa e uma função de verossimilhança, em nível de marcas individuais, possa explicar alguma informação adicional anteriormente não detectada. Diante do exposto, este estudo visou propor o método Bayes Híbrido e compará-lo ao BLASSO quanto à eficiência na estimação dos valores genômicos utilizando dados simulados para dois cenários com herdabilidade em torno de 0,20, com ausência de dominância e com dominância completa. Ambos os métodos, apresentaram

---

<sup>1</sup> Programa de Pós-graduação em Estatística Aplicada e Biometria – UFV. e-mail [leisa.lima@ufv.br](mailto:leisa.lima@ufv.br)

<sup>2</sup> Departamento de Estatística – UFV. e-mail [camila.azevedo@ufv.br](mailto:camila.azevedo@ufv.br)

<sup>3</sup> Embrapa Floretas/Departamento de Engenharia Florestal. e-mail [marcos.deon@gmail.com](mailto:marcos.deon@gmail.com)

<sup>4</sup> Departamento de Zootecnia – UFV. e-mail [fabyanofonseca@ufv.br](mailto:fabyanofonseca@ufv.br)

<sup>5</sup> Departamento de Biologia Geral – UFV. e-mail [jmsviana@ufv.br](mailto:jmsviana@ufv.br)

<sup>6</sup> Agradecimento a CAPES, a FAPEMIG e a Fundação Arthur Bernardes (FUNARBE) pelo apoio financeiro.

valores de herdabilidades e acurácias similares. Isso pode ocorrer, pois o BLASSO é assintoticamente livre de informação *a priori*, ou seja, proporciona uma melhor aprendizagem por meio do conjunto de dados. Entretanto, o Bayes Híbrido conduziu as estimativas de valores genômicos com viés mais próximo de zero e assim, pode ser utilizado vantajosamente na GWS.

**Palavras-chave:** *Índice, BLASSO, BLUP genômico.*

## **Predição Genômica de Caracteres Quantitativos por meio de Redes Neurais Artificiais após redução de dimensionalidade.**

**Isabela de Castro Sant' Anna<sup>1 4</sup>**

**Weverton Ribeiro Gomes<sup>2 4</sup>**

**Gabi Nunes Silva<sup>2 4</sup>**

**Moysés Nascimento<sup>2 4</sup>**

**Cosme Damião Cruz<sup>3 4</sup>**

Um dos grandes desafios do melhoramento genético atual é entender a variação da herança genética de caracteres quantitativos que são condicionados por muitos genes com pequeno efeito cuja interação resulta em não linearidade entre fenótipos e genótipos. A inclusão dessas interações conduz à superparametrização dos modelos de seleção genômica (SG) devido ao número de marcadores do genoma ser maior que o número de indivíduos genotipados. Assim, as Redes Neurais Artificiais (RNAs) tornam-se alternativas eficientes por capturar relações não lineares entre marcadores a partir de fenótipos. Entretanto, a inclusão dos marcadores no modelo aumenta a dificuldade de aprendizado da RNA. Assim, um modelo mais realístico incluiria apenas SNPs relacionados a característica de interesse. Consequentemente, foi proposto o uso de métodos de redução de dimensionalidade (RD), para fins de seleção de um subconjunto de marcadores por meio de procedimentos específicos como Sonda ou regressões *Stepwise*. Com o objetivo de avaliar a eficiência da seleção SG e RNA na predição do valor genético em população natural com desequilíbrio gamético. Para isso, foi simulada uma população F1, genotipados com 1000 SNP. As características fenotípicas foram determinadas adotando-se diferentes cenários com  $h^2$  30%. O quadrado da correlação entre o valor genômico predito e o valor fenotípico foi utilizado para medir a confiabilidade preditiva. No cenário aditivo, o  $R^2$  de validação foi de 59% para RBF, e 57% para RR-BLUP, e no cenário epistático

---

<sup>1</sup> DET– UFV e-mail isabelacsantanna@gmail.com

<sup>2</sup> DET– UFV

<sup>3</sup> DBG– UFV

<sup>4</sup> Agradecimento a Capes, FAPEMIG e CNPq pelo apoio financeiro

os valores de  $R^2$  foram de 50%, 47 % respectivamente. OS resultados indicam que a RD melhora as acurácias em todos os cenários utilizados.

**Palavras-chave:** *Espistasia, Aprendizado de maquinas, seleção genômica.*

## Procedimentos biométricos na avaliação de efeitos epistáticos

Alexandre Gomes Ferraz<sup>1 3</sup>

Cosme Damião Cruz<sup>1 3</sup>

Isabela de Castro Sant'Anna<sup>2 3</sup>

Weverton Gomes da Costa<sup>1 3</sup>

Iara Gonçalves dos Santos<sup>1 3</sup>

O trabalho objetiva entender como a análise dialélica de Hayman detecta desvio da pressuposição do modelo aditivo dominante e, na presença da epistasia, gerar informações acuradas. Nas análises usaram-se dados simulados, onde o efeito epistático sobre a característica era conhecido. Simularam-se os dados adotando-se três modelos: modelo aditivo, aditivo-dominante e epistático, em três situações: ausência de dominância, dominância parcial e dominância completa. Foram simulados caracteres quantitativos com herdabilidades ( $h^2$ ) de 30 e 60% e controlados cada um por 20 locos, sendo dois alelos por loco. O teste de suficiência do modelo aditivo-dominante foi feito por meio da regressão (com e sem rotação) dos valores da covariância entre valores médios dos genitores e de sua linha ( $W_r$ ) em função da variância da linha ( $V_r$ ). No teste de regressão, sem rotação, característica de alta  $h^2$  (60%), os resultados foram 100% concordantes com o esperado rejeitando-se o modelo aditivo-dominante quanto havia ação epistática no controle gênico da característica e não rejeitando quanto a epistasia estava ausente. Quando  $h^2$  da característica era baixa (30%), a regressão com rotação adequou-se para demonstrar ocorrência de falhas no modelo aditivo-dominante na presença de epistasia. Sem rotação, o teste apontou falhas nos cenários em que a epistasia não estava presente no controle gênico gerando resultados falsos positivos. Conclui que a técnica biométrica de análise dialélica de Hayman é um procedimento eficaz na mensuração de efeitos de interação gênica, principalmente, em característica de  $h^2$  elevada. Efeitos ambientais elevados podem mascarar a detecção de ação gênica proporcionada por interações interalélicas.

**Palavras-chave:** *Melhoramento, biometria, genética.*

<sup>1</sup> DBG - UFV. e-mail: [alexandre.g.ferraz@gmail.com](mailto:alexandre.g.ferraz@gmail.com)

<sup>2</sup> DET - UFV.

<sup>3</sup> Agradecimentos: FAPEMIG, CNPq, Capes, UFV



## Seleção de Variáveis via método BLASSO

Roberta de Amorim Ferreira<sup>1 3</sup>

Luiz Alexandre Peternelli<sup>2 3</sup>

Muitos conjuntos de dados apresentam um grande volume de variáveis (alta dimensionalidade) a serem analisadas. Assim, métodos específicos para ajuste do modelo devem ser empregados. Embora os métodos existentes sejam, em geral, eficientes para construção do modelo, frequentemente se faz necessário selecionar as variáveis mais explicativas do modelo. O método BLASSO, além de modelar os dados, pode funcionar como um método de seleção, visto que faz alguns coeficientes tenderem a valores próximos a zero e então um limite de significância pode ser estabelecido de forma que os coeficientes de regressão abaixo desse limite sejam descartados. Esse trabalho propõe avaliar o BLASSO sob alguns critérios de seleção em que 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50% das variáveis mais significativas são selecionadas, e avaliar a melhoria na qualidade de predição, quando comparado ao uso dos dados completos. O teor de fibra de 168 genótipos de cana-de-açúcar, obtidos do programa de melhoramento da UFV foram dispostos em uma matriz com 168 linhas e 3113 colunas. A qualidade preditiva foi avaliada baseada no coeficiente de correlação entre valores preditos e reais ( $r$ ) e na raiz quadrada do erro quadrático médio (RMSE). No modelo completo, os valores encontrados foram  $RMSE = 1,74$  e  $r = 0,685$  para predição. Com o critério de seleção o resultado que mais se aproximou do modelo completo foi obtido com 25% das variáveis, com  $RMSE = 1,73$  e  $r = 0,687$ . A seleção de variáveis pouco melhorou a qualidade preditiva do modelo.

**Palavras-chave:** *Alta dimensionalidade, cana-de-açúcar, bayesian lasso.*

---

<sup>1</sup> DET – UFV. e-mail: roberta.amorim@ufv.br

<sup>2</sup> DET – UFV

<sup>3</sup> Agradecimentos à RIDESA, à CAPES, à FAPEMIG e ao CNPq pelo apoio financeiro.

## Uso de *B-spline* para estimar a função sinal no modelo genoma contínuo

Ernandes Guedes Moura <sup>1 4</sup>  
Andrezza Kellen Alves Pamplona <sup>2 4</sup>  
Marcio Balestre <sup>3 4</sup>  
Júlio Sílvio de Sousa Bueno Filho <sup>3 4</sup>

O uso do modelo genoma contínuo de Hu et al. (2012) na seleção genômica consiste na ideia de que a expressão gênica é uma função da posição no genoma. Como o objetivo é estimar tal função, técnicas de ajuste de curvas polinomiais com baixa ordem parecem fazer bem: o intervalo de interesse é subdividido em partes menores e a função de interesse pode ser aproximada por polinômios definidos localmente de grau relativamente baixo. Com isso, o presente estudo tem como objetivo utilizar modelos funcionais e estimar a função sinal por meio do sistema de bases *B-spline*. Além disso, busca-se verificar se tal abordagem apresenta vantagens preditivas com relação aos métodos RR-BLUP e Bayes A. Um genoma com 10010 marcadores foi simulado e os métodos foram analisados nos cenários oligogênico e poligênico para as herdabilidades 0,3 e 0,7. Adotou-se um polinômio de grau dois para ajustar a *B-spline*, pois se supõe que entre dois nós adjacentes não existe mais que um QTL com efeito expressivo. No cenário oligogênico, o método RR-BLUP foi considerado o mais acurado nas duas herdabilidades; já no cenário poligênico, a capacidade preditiva de *B-spline* foi análoga à do RR-BLUP, ambos sendo os mais acurados nas duas herdabilidades. O método *B-Spline* foi o que apresentou menor custo computacional. Nosso estudo mostrou que a abordagem *B-spline* pode lidar com um número ilimitado de marcadores e com custo computacional muito menor do que os modelos tradicionais avaliados.

**Palavras-chave:** *Spline, B-spline, Modelo genoma contínuo, Valor genômico.*

---

<sup>1</sup>IFMA - campus São João dos Patos. Doutorando DES - UFLA.

<sup>2</sup>IFTM - campus Uberaba. Doutoranda DES - UFLA. e-mail: [andrezzapamplona@iftm.edu.br](mailto:andrezzapamplona@iftm.edu.br)

<sup>3</sup>DES - UFLA.

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## Variabilidade genética no programa de melhoramento genético de arroz irrigado

Weverton Gomes da Costa<sup>1 3</sup>

Antônio Carlos da Silva Júnior<sup>2 3</sup>

Iara Gonçalves dos Santos<sup>2 3</sup>

Isabela de Castro Sant'anna<sup>2 3</sup>

Cosme Damião Cruz<sup>2 3</sup>

É a variabilidade genética da espécie que determina o seu conjunto de características morfológicas e fisiológicas, o que a torna capaz de responder às mudanças ambientais. Visto a importância deste tema, o presente trabalho visa estudar a variabilidade genética do programa de melhoramento genético de arroz irrigado da EPAMIG em Minas Gerais a partir das estimativas dos parâmetros genéticos das linhagens. Para realização deste estudo utilizou-se os dados de produtividade de grãos e altura de plantas avaliados nas safras de 2004/2005, 2008/2009, 2011/2012 e 2015/16 testados nas localidades de Leopoldina-MG e Lambari-MG. Foi realizada Análise de Variância individual e as médias agrupadas pelo teste Scott-Knott, onde cada localidade dentro de cada safra foi considerada como um ambiente independente. A variabilidade genética foi verificada pelo método de otimização de Tocher, a partir da distância de Mahalanobis. Foram detectadas diferenças significativas para cada ambiente para as duas características estudadas. O coeficiente de variação (CV) para produtividade de grãos variou de 10,9% a 21,5% e para altura de plantas de 1,5% a 6,7%. A herdabilidade foi o parâmetro genético que mais oscilou durante os anos. A menor herdabilidade foi encontrada em Lambari na safra de 2004/2005 e a maior na mesma safra, porém em Leopoldina. Houve formação de 4 grupos de ambientes pelo método de otimização de Tocher, demonstrando que existe variabilidade genética dentro do programa ao longo dos anos. A existência de variabilidade genética aponta a possibilidade de obterem-se ganhos genéticos com a seleção para todos os caracteres.

**Palavras-chave:** *Oryza sativa* L., Tocher, Variabilidade.

<sup>1</sup> DBG – UFV. e-mail wevertonufv@gmail.com

<sup>2</sup> DBG – UFV.

<sup>3</sup> Agradecimento ao CNPq e à FAPEMIG pelo apoio financeiro.

## Variáveis associadas a qualidade de cafés especiais de Minas Gerais

Alberto Frank Lázaro Aguirre <sup>1 3</sup>

Eric Batista Ferreira <sup>2 3</sup>

Júnio César Rosa <sup>1 3</sup>

Luiz Otávio de Oliveira Pala <sup>1 3</sup>

A conjuntura econômica de mercados cada vez mais competitivos tem motivado consumidores quanto às variedades disponíveis de café no mercado, influenciando produtores a optarem por formas de aprimoramento quanto a qualidade do grão e, expandindo a produção de cafés especiais com o objetivo do aumento da demanda por este produto. Nesta problemática, este trabalho teve por objetivo a detecção de variáveis comuns às amostras vencedoras no Concurso Mineiro de Qualidade do Café, realizado no município de Lavras no ano de 2013, que classificou as amostras quanto aos níveis de ácidos graxos: palmítico, esteárico, oleico, linoleico, linolênico, araquídico e behênico; variáveis físicas: trigonelina, ácido clorogênico, cafeína e sacarose; e sensoriais: acidez, corpo, sabor e doçura. Para tanto, utilizou-se da metodologia de componentes principais e análise de agrupamentos, explorando padrões e variáveis relacionadas com às amostras vencedoras do concurso. Como resultado, a análise de agrupamentos evidenciou grupos entre as amostras, aglomerando as melhores colocadas e possibilitando a interpretação dos componentes principais, indicando que as amostras vencedoras estiveram associadas aos maiores níveis de ácido linoleico e menores de sacarose.

**Palavras-chave:** *Café, Qualidade do café, Componentes principais.*

---

<sup>1</sup>Mestrando em Estatística Aplicada e Biometria - UNIFAL/MG. e-mail: flazaro2857@gmail.com; juninhocapitolio@gmail.com; luizotavio.oliveira@gmail.com.

<sup>2</sup>Professor Associado I do Instituto de Ciências Exatas da Universidade Federal de Alfenas.

<sup>3</sup>Agradecimento ao CNPq, FAPEMIG, CAPES e UNIFAL pelo apoio financeiro.

## WAAS: um novo índice de estabilidade em análise AMMI

Tiago Olivoto <sup>1 3</sup>

Alessandro Dal'Col Lúcio <sup>2 3</sup>

Bruno Giacomini Sari <sup>2 3</sup>

Maria Inês Diel <sup>2 3</sup>

Dionatan Ketzer Krysczun <sup>2 3</sup>

Em modelos AMMI, a interpretação da interação genótipo-*vs*-ambiente (GA) é baseada, geralmente, na dispersão dos escores dos dois primeiros componentes principais (CP). Em algumas situações, a soma de quadrados da interação GA (SQGA) capturada nos dois primeiros CPs não é suficientemente alta. Aqui, demonstramos como este problema pode ser contornado. Dois conjuntos de dados foram utilizados: 1) 29 cultivares de soja cultivadas em quatro ambientes; 2) 10 cultivares de aveia cultivadas em 16 ambientes. Diferentemente do *biplot* tradicional (RG × CP1), um novo índice foi utilizado, denominado como média ponderada dos escores absolutos (WAAS), estimada por:  $WAAS_i = \sum_{k=1}^k |CP_{ik} \times PE_k| / \sum_{k=1}^k PE_k$ , onde  $CP_{ik}$  e  $PE_k$  são, respectivamente, o escore do  $i$ -ésimo genótipo no  $k$ -ésimo CP e a percentagem da SQGA explicada pelo  $CP_i$ . O genótipo com o menor valor de WAAS pode considerado o mais estável. Com o objetivo de atribuir diferentes pesos ( $P$ ) para produtividade ( $P$ ) e estabilidade ( $E$ ) na classificação dos genótipos, a WAASY também foi estimada por:

$$WAASY_i = \frac{\left\{ P_p \times \left[ \left( \frac{GY_i}{GY_{\max}} \right) \times 100 \right] \right\} + \left\{ P_e \times \left( 100 - \frac{WAAS_i}{WAAS_{\min}} \right) \right\}}{P_p + P_e}. \text{ Diferentes combinações de pesos (21) foram}$$

estudadas. Comparado com o *biplot* tradicional, foi possível capturar 30,7% e 60% a mais da SQGA nos experimentos 1 e 2, respectivamente, utilizando a WAAS. A correlação CP1×WAAS obtida com os ranques das cultivares foi de 0.75, demonstrando que a maior captura da SQGA alterou o ranqueamento das cultivares quanto a estabilidade. Utilizando gráficos heatmap foi possível identificar quatro grupos de genótipos: produtivos e estáveis, produtivos mas não estáveis, estáveis mas não produtivos e não estáveis e não produtivos.

**Palavras-chave:** Interação, Escores, WAASY

<sup>1</sup> Departamento de Fitotecnia – UFSM. e-mail: tiagoolivoto@gmail.com

<sup>2</sup> Departamento de Fitotecnia – UFSM

<sup>3</sup> Agradecimentos à CAPES e CNPq pelo suporte financeiro

## Meio Ambiente

## A Estatística Circular como ferramenta de análise da direção do vento em Santa Maria, RS

Augusto Maciel da Silva <sup>1 4</sup>

Enio Junior Seidel <sup>2 4</sup>

Helen da Silva Costa Benaduce <sup>3 4</sup>

O conhecimento da direção dos ventos em determinada região é de suma importância para construção civil, planejamento agrícola, controle de poluição, entre outros. A direção do vento é baseada em sua origem, podendo ser representada por um vetor com direção e sentido que forma um ângulo com o Norte ( $0^\circ$ ). Dessa forma, um conjunto de dados mensurados em graus, como a direção do vento, deve ser tratado de forma diferente de dados lineares, devido a sua periodicidade e disposição circular. A Estatística Circular fornece as ferramentas adequadas para tratar dados angulares. O presente trabalho tem por objetivo a construção de diagramas circulares para análise da uniformidade, obtenção de medidas descritivas, momentos trigonométricos e adequação a Distribuição Circular Uniforme da direção dos ventos no período de setembro de 2016 a agosto de 2017 na cidade de Santa Maria/RS. Os dados foram disponibilizados pelo Instituto Nacional de Meteorologia ? INMET. Como resultado, verificou-se tanto graficamente quanto através do Teste de Watson, a ausência de uniformidade na direção dos ventos, tanto para o período anual como também quando separados por estações do ano. A média circular anual foi de  $101^\circ$ , com variância circular de 0,58, caracterizando ventos concentrados na direção Leste e Sudeste. Considerando as quatro estações, verificou-se comportamento diferente no Inverno, com ventos concentrados na direção média de  $77^\circ$  (Leste e Nordeste) Com o conhecimento dessas características e da não uniformidade dos ventos, tem-se um melhor planejamento para setores que sofram influência direta dos ventos.

**Palavras-chave:** *Direção dos Ventos, Distribuição Circular Uniforme, Estatística Circular.*

---

<sup>1</sup>Departamento de Estatística - UFSM. e-mail: [augusto.silva@ufsm.br](mailto:augusto.silva@ufsm.br)

<sup>2</sup>Departamento de Estatística - UFSM. e-mail: [enioseidel@gmail.com](mailto:enioseidel@gmail.com)

<sup>3</sup>Bacharelado em Estatística - UFSM

<sup>4</sup>Thanks to CNPq for the financial support.

## Análise de dados experimentais com resposta binária e elevadas e/ou baixas proporções de sucesso

Luma dos Passos Bispo <sup>1</sup>

Maurício Santana Lordêlo <sup>2</sup>

Gilênio Borges Fernandes <sup>3</sup>

Em dados experimentais com resposta binária é frequente o pesquisador usar pequeno número de tentativas e ocorrer baixas e/ou elevadas proporções de sucesso nos tratamentos (grupos). Nestes casos, sabe-se que a variância não é constante à medida que a média muda. A análise é feita geralmente através dos modelos logit, probit, complemento log-log, etc. Neste cenário, se a forma assumida para a relação entre a resposta média e os preditores estiver correta, as estimativas dos parâmetros do preditor linear por máxima verossimilhança ou mínimos quadrados não linear são imparciais, mas, as estimativas da matriz de covariâncias das estimativas dos parâmetros é viesada, prejudicando as inferências. Uma estratégia é usar estimadores robustos (heteroscedasticidade consistente, HC, tipo sandwich) para a matriz de covariância. Estimadores tipo sandwich foram desenvolvidos para o modelo de regressão linear clássica e estendidos para o preditor linear dos modelos lineares generalizados. Resultados preliminares numa aplicação com seis grupos onde as proporções de sucesso variaram de zero a cem por cento, o estimador sandwich mostrou melhoria expressiva nas inferências em relação ao método trivial. Para comparações múltiplas de proporções, sugere-se testes de hipóteses em combinações lineares de parâmetros e restrições lógicas. O objetivo deste estudo é comparar, por meio de dados simulados e reais, a eficiência do estimador tipo sandwich com o convencional na análise de dados experimentais com resposta binária com elevadas e/ou baixas proporções de sucesso.

**Palavras-chave:** *dados binários, estimador sandwich, heterocedasticidades.*

---

<sup>1</sup>Doutoranda PPGRGV - UEFS. e-mail: [luma.pb@hotmail.com](mailto:luma.pb@hotmail.com)

<sup>2</sup>DEXA - UEFS. e-mail: [mslordelo@uefs.br](mailto:mslordelo@uefs.br)

<sup>3</sup>DEST - UFBA, DEXA - UEFS. e-mail: [gileniofb@gmail.com](mailto:gileniofb@gmail.com)



## Aplicação do modelo logístico na estimação da probabilidade de se encontrar depósitos de ouro na Índia

Vinicius Garcia Tosta <sup>1 3</sup>

Everton da Costa <sup>2 3</sup>

Nesse presente trabalho, foi aplicado um modelo logístico para se estimar a probabilidade de ocorrência de depósitos de ouro em uma parte do cinturão de xisto aurífero Arqueano Hutti-Maski, na Índia. Os dados utilizados consistem em variáveis categóricas e contínuas obtidas a partir de um mapa de linha codificado e mapas de anomalia geoquímica dos elementos presentes no caminho do ouro tanto no solo quanto em águas subterrâneas.

Antes de se construir o modelo logístico propriamente dito, os dados foram divididos do seguinte modo: 70% foram utilizados para a construção do modelo e os 30% restantes foram utilizados na validação por intermédio da curva ROC.

A variável resposta “ouro” é do tipo binária, recebendo 1 para presença de depósito de ouro e 0 para ausência. As variáveis preditoras são as concentração de arsênio e de antimônio no solo, bem como a presença ou não de lineamento (1 ou 0).

De acordo com o modelo logístico, apenas a concentração de arsênio foi significativa, ao passo que o mesmo não ocorreu para as outras duas variáveis. Assim sendo, a cada aumento de uma unidade na concentração de arsênio, a chance de se encontrar um depósito de ouro aumenta em 3.34 vezes.

De acordo com a curva ROC, o modelo obteve simplesmente uma capacidade preditiva perfeita, alcançando 100% tanto para a sensibilidade quanto para a especificidade. Assim, a concentração de arsênio parece ser um forte indício para se encontrar depósitos de ouro nessa região.

**Palavras-chave:** *Ouro, Hutti-Maski, Curva ROC.*

---

<sup>1</sup>DES - UEM. e-mail: [tostavinicius@yahoo.com.br](mailto:tostavinicius@yahoo.com.br)

<sup>2</sup>DES - UEM.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

## Avaliação de modelos para a determinação do diâmetro por meio de anéis de crescimento

Douglas Rufino Vaz <sup>1 3</sup>

Saimom Poczapski Noro Ribeiro <sup>1 3</sup>

Crysttian Arantes Paixão <sup>2 3 4</sup>

Heloisa Maria de Oliveira <sup>2 3 4</sup>

Mário Dobner Junior <sup>2 3 5</sup>

**Resumo:** Considerado um dos aspectos de maior importância para avaliação de uma espécie florestal, os anéis anuais de crescimento são capazes de registrar inúmeras características de significativa importância para a árvore. Sua capacidade de manter registrado inúmeros fenômenos edafoclimáticos, proporcionaram a criação de uma nova ciência, conhecida como Dendrocronologia. Tendo por objetivo mensurar os anéis, foi selecionado um disco de uma árvore de *Pinus taeda*, com 30 anos de idade, localizada no município de Campo Belo do Sul, Santa Catarina, Brasil. Com esta finalidade, inúmeros modelos matemáticos foram analisados, considerando os dados obtidos do disco, destacando-se: Horel, Backman, Gram e Moissev. Após as análises estatísticas, o modelo de Horel foi considerado o de melhor ajuste, indicado para avaliar o crescimento desta árvore, permitindo realizar as previsões até os 70 anos.

**Palavras-chave:** anéis de crescimento; diâmetro; modelo de Horel; modelos matemáticos; *Pinus taeda*.

---

<sup>1</sup>CRC - Estudante de Engenharia Florestal do Centro de Ciências Rurais, Centro de Curitibaanos.  
e-mail: [douglasrufinovaz@gmail.com](mailto:douglasrufinovaz@gmail.com)

<sup>2</sup>CRC - Professor(a) Dr(a). do Centro de Ciências Rurais, Centro de Curitibaanos).

<sup>3</sup>UFSC - Universidade Federal de Santa Catarina, Curitibaanos, SC

<sup>4</sup>CNS - Departamento de Ciências Naturais e Sociais

<sup>5</sup>ABF - Departamento de Agricultura, Biodiversidade e Florestas

## **Comparação de modelos estatísticos para estimação do intervalo de tempos entre ultrapasses de um limiar de temperatura na cidade de Presidente Prudente (SP)**

**Maria Magdalena Kcala Alvaro<sup>1</sup>**

**Mário Hissamitsu Tarumoto<sup>2</sup>**

**Eliane Regina Rodrigues<sup>3</sup>**

A observação de fenômenos naturais, como a chuva e temperatura é bastante frequente no mundo atual, de forma que vários estudos têm sido realizados com o intuito de prever a ocorrência delas tendo em vista o que ocorreu no passado. Dados desta natureza, em que a coleta de dados ocorre de forma contínua, seja por medida horária ou diária, não apresenta independência entre as observações. Entre as possíveis formas de análise, poderia ser por meio da aplicação de técnicas de séries temporais, no entanto, o objetivo deste estudo é construir uma matriz de transição, de tal forma que possamos determinar a probabilidade, por exemplo, de alta temperatura amanhã, dado que hoje foi observado este fenômeno. Para a obtenção deste resultado, uma possibilidade é construir um modelo baseado em dados dependentes, como por exemplo, um processo de markov, em que a suposição é de que exista dependência somente com o dia anterior. O objetivo deste trabalho foi a de construir este modelo e realizar a aplicação em dados de temperatura na cidade de Presidente Prudente no período de janeiro de 2005 a dezembro de 2017. Foi construído a verossimilhança condicionada para a distribuição Weibull por meio da utilização de funções cópulas.

**Palavras-chave:** *Cópulas, tempos entre ultrapasses, processo de markov.*

---

<sup>1</sup> Depto de Estatística – FCT/UNESP – Prog.PG em Mat. Aplicada e Computacional

<sup>2</sup> Depto de Estatística – FCT/UNESP – Prog.PG em Mat. Aplicada e Computacional – e-mail: mario.tarumoto@eduapps.unesp.br

<sup>3</sup> Instituto de Matemática – Universidad Nacional Autónoma de México.

## Comparação dos modelos não lineares de isotermas de adsorção de potássio no Ceará.

Kelly Pereira de Lima <sup>1 3</sup>

Marcela de Marillac Carvalho <sup>2 3</sup>

Jaqueline Gonçalves Fernandes <sup>2 3</sup>

Elias Silva de Medeiros <sup>2 3</sup>

Augusto Ramalho de Moraes <sup>2 3</sup>

A adsorção é um fenômeno que pode ser descrito por isotermas de adsorção, que está relacionada diretamente com o volume adsorvido em função da pressão ou concentração do adsorvente com temperatura constante. Existem algumas isotermas de adsorção, como por exemplo, as isotermas de Langmuir, Freundlich, Redlich-Peterson, Fritz, Sips e etc. No caso, as isotermas são modelos não lineares, mas são facilmente linearizadas, porém a linearização pode alterar a estrutura e distribuição do erro, assim, pode ocorrer que os erros do modelo original satisfaçam as pressuposições usuais de normalidade, independência e homogeneidade da variância, já os erros do modelo linearizado podem não satisfazê-las. Com isso, esse trabalho tem como objetivo comparar os modelos não lineares de isotermas. Os dados são provenientes de Filho (2012) que se referem adsorção de equilíbrio do potássio na profundidade de 0–2.5 cm na área de mata nativa no Ceará, no qual, têm-se observações para a concentração de equilíbrio ( $\text{mg } L^{-1}$ ) e para o potássio adsorvido ( $\text{mg } kg^{-1}$ ). Os modelos Freundlich e Redlich-Peterson foram ajustados aos dados experimentais. Além disso, para verificar os pressupostos de normalidade, homogeneidade e independência dos resíduos foram utilizados os testes de Shapiro Wilk, Breusch Pagan e Durbin Watson, respectivamente. A qualidade do ajuste dos modelos foi feita com base nos valores do coeficiente de determinação ajustado e do critério de informação de Akaike (AIC). Conclui-se que o modelo Redlich-Peterson apresentou o menor valor de AIC e maior valor de coeficiente de determinação ajustado, assim sendo, o melhor ajuste para os dados.

**Palavras-chave:** *Química, Curvas, Concentração.*

---

<sup>1</sup>DES -UFLA. e-mail: *kelly.lima.88@gmail.com*

<sup>2</sup>DES - UFLA.

<sup>2</sup>DES - UFLA.

<sup>2</sup>DES - UFLA.

<sup>2</sup>DES - UFLA.

<sup>3</sup>Agradecimento ao CNPq, Capes e FAPMIG pelo apoio financeiro.

## Diagnóstico e Análise Estatística da Precipitação das Estações Meteorológicas da Bahia

Eucymara França Nunes Santos<sup>1 3</sup>

Inajá Francisco de Souza<sup>2</sup>

Íkaro Daniel de Carvalho Barreto<sup>3</sup>

**Resumo Simples:** A climatologia tem sido objeto de pesquisa pela Meteorologia com o auxílio das ferramentas Estatísticas, então a proposta deste artigo é utilizar a técnica de análise de regressão múltipla aplicada a previsão do tempo, para diagnóstico e previsão da variável precipitação pluviométrica, para descrever e prever o comportamento dos dados meteorológicos com o intuito de planejar para diminuir os riscos das atividades agrícolas. A modelagem estatística da variável resposta precipitação em relação às médias mensais é aplicada em 20 estações meteorológicas do Estado da Bahia, considerando 30 anos de observações, nos 12 meses do ano, anual e períodos seco e chuvoso. Verifica-se um bom desempenho dos modelos, onde todos os valores apresentados estão próximos de zero, indicando menor erro. Os valores médios que representam o período com maior intensidade de precipitação estão entre os meses de Novembro a Abril. O melhor modelo apresentado foi para o mês de novembro, com 72,56% de associação entre as variáveis.

Palavras-chave: *Estações Meteorológicas, Precipitação e Regressão Múltipla*

---

<sup>1</sup> Professora Mestre do Departamento de Estatística e Ciências Atuariais, Aracaju, Universidade Federal de Sergipe. Doutoranda do PPGBEA - UFRPE, [eucymara@gmail.com](mailto:eucymara@gmail.com)

<sup>2</sup> Professor Doutor da Pós Graduação em Recursos Hídricos, Aracaju, Universidade Federal de Sergipe, [inajafrancisco@gmail.com](mailto:inajafrancisco@gmail.com)

<sup>2</sup> Doutorando da Universidade Federal Rural de Pernambuco – PPGBEA – UFRPE, [daniel.carvalho.ib@gmail.com](mailto:daniel.carvalho.ib@gmail.com)

<sup>3</sup> Agradecimento à CAPES pelo apoio financeiro.

## Fatores que influenciam a produção de cerne e de alburno em teca

Édila Cristina de Souza<sup>1 3</sup>

Fernanda Viana da Silva Leonardo<sup>2 3</sup>

Diego Tyszka Martinez<sup>3 3</sup>

A *Tectona grandis* é uma espécie arbórea de grande porte, valorizada economicamente devido a beleza na produção de móveis de luxo e na construção naval. A oferta de matéria-prima proveniente de reflorestamentos desperta preocupação de madeireiros com a qualidade. Na silvicultura, as decisões de manejo são baseadas na produção quantitativa/volumétrica. O objetivo foi avaliar as características de cerne e alburno e as interferência de fatores(ambientais/silviculturais). O experimento foi em Figueirópolis D'Oeste(MT), com clones da Malásia de oito anos. Os fatores foram: material de origem (clonal e seminal); face de exposição à radiação solar(norte, sul, leste, oeste); posição sociológica(intermediária, suprimida e oprimida); circunferência(1,3 m do solo); a altura total; raios de projeção de copa. Foram calculados: diâmetro a 1,3 m do solo(DAP), altura dominante, área de seção transversal, a área basal, volume total/hectare, quantificação da área de cerne e de alburno e as proporções destas na seção transversal(DAP). As variáveis foram submetidas a Análise de Componentes Principais(ACP), Análise Fatorial(AF) e Análise de Variância Multivariada(MANOVA) e Tukey. O crescimento da teca de origem clonal foi superior, mesmo submetida a estresse causado pelo vento. A posição sociológica interfere nas características do lenho em teca e o cerne e alburno são influenciados por fatores genéticos.

**Palavras-chave:** *Alburno, Cerne, Multivariada.*

<sup>1</sup> DEST – UFMT. edilacrsouza@gmail.com

<sup>2</sup> FENF – UFMT. viana.fernandaleonardo@gmail.com

<sup>3</sup> FENF – UFMT. diegotyszka@hotmail.com

# Modeling the emission of $CO_2$ through the cement production in Brazil between the years 1901 and 2014 using the ARFIMA model

Aline Edlaine de Medeiros <sup>1 2</sup>

MÃ ´nica Marin de Souza <sup>3</sup>

Eniuce Menezes de Souza <sup>4</sup>

Edson Zangiacomi Martinez <sup>5</sup>

We employ the ARFIMA model to adjust the  $CO_2$  emissions related to the annual average of the cement production in Brazil between the years 1901 and 2014. The use of this model was driven the fact that this time series has long-term persistent behavior.

**Keywords:** *Autocorrelation, ARFIMA, Time Series.*

---

<sup>1</sup>PBE - UEM. e-mail: [aline\\_medeiros@outlook.com](mailto:aline_medeiros@outlook.com)

<sup>2</sup>Agradecimento ao CAPES pelo apoio financeiro.

<sup>3</sup>PBE - UEM.

<sup>4</sup>DES - UEM

<sup>5</sup>FMRP - USP

## Modelagem do número de autos de infrações ambientais no estado de Minas Gerais para os anos de 2008 a 2016

Marcel Irving Pereira Mélo<sup>1 4</sup>

Sarita Soraia de Alcântara Laudares<sup>2 4</sup>

Júlio Sílvio de Sousa Bueno Filho<sup>3 4</sup>

Ao longo de toda história da legislação ambiental é possível observar reincidências na reformulação das leis, a fim de se fazerem cumpridas, fato observado, inclusive, sobre a Lei de Proteção da Vegetação Nativa (LPVN nº 12.651 de 2012), popularmente conhecida como Novo Código Florestal. Esta Lei, sancionada com intuito de sanar o passivo acumulado por anos de descumprimento do antigo Código Florestal (Lei nº 4.771/1965) por meio da intervenção em áreas protegidas. Isso posto, temos o seguinte questionamento: será que após a sanção do Novo Código Florestal a tutela jurídica passou a ser respeitada? Para tanto, foram avaliados no estado de Minas Gerais 4518 autos de infração obtidos pelo Sistema de controle de Autos de Infração e Processos (CAP) da Secretaria de Estado de Meio-Ambiente e Desenvolvimento Sustentável - SEMAD, entre os anos de 2008 a 2016. As autuações foram agrupadas por ano e mesorregião, dessa forma, foi construído um modelo latente gaussiano para o número de autuações, para a qual foi assumido distribuição Poisson com função de ligação logarítmica entre a média e o preditor. As variáveis explicativas foram o ano da autuação, mesorregião da autuação e interação mesorregião e ano, para os quais foram assumidas prioris gaussianas. Os resultados apontam uma tendência crescente ao longo dos anos para o número de autuações e uma forte dependência espacial na mesorregião metropolitana de Belo Horizonte, sugerindo que a implantação e sanção do Novo Código Florestal vem se mostrando ineficaz para a redução dos autos de infrações ambientais.

**Palavras-chave:** *Integrated Nested Laplace Approximation (INLA), Modelo Autorregressivo Condicional (CAR), Novo Código Florestal.*

---

<sup>1</sup>Departamento de Estatística - UFLA. e-mail: [mipmelo90@gmail.com](mailto:mipmelo90@gmail.com)

<sup>2</sup>Departamento de Engenharia Florestal - UFLA. email: [saritalaudares@gmail.com](mailto:saritalaudares@gmail.com)

<sup>3</sup>Departamento de Estatística - UFLA. e-mail: [jssbueno@des.ufla.br](mailto:jssbueno@des.ufla.br)

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.



## Perfil da Qualidade da Água por meio dos Parâmetros Físicos-Químicos-Biológicos e Indicador no Estado do Paraná

Gabriel Sartori Klostermann <sup>1</sup> Paulo Justiniano Ribeiro Junior <sup>2</sup>

A bacia hidrográfica pode ser caracterizada por um conjunto de terras drenadas por um rio e seus afluentes, formada nas regiões mais altas do relevo por divisores de água, onde as águas das chuvas, escoam superficialmente formando os riachos e rios. Além da questão natural da bacia, é também uma unidade de gestão de recursos hídricos, no qual comitês de Bacias, buscam promover o planejamento de controlar os usos da água na região, proteger e conservar as fontes de captação da bacia, portanto, se faz necessário avaliar a qualidade da água. O objetivo geral do trabalho é analisar, comparar e diagnosticar a evolução dos parâmetros coletados na água, e do IQA (índice de qualidade da água) que resume nove parâmetros, a nível de bacia hidrográfica e nas estações de monitoramento. O banco de Dados é composto por coletas do instituto AguasParaná e pelo Professor Cristóvão do DHS/UFPR, no qual registra-se, cerca de 8500 coletas em 13 bacias hidrográficas de 294 estações de monitoramento no período de 1982 à 2015. As análises mostram que o IQA variou ao longo do tempo, com um comportamento próximo entre as bacias e que o principal parâmetro responsável por esta oscilação foi o coliformes fecais. Além disso, verificou-se que a qualidade esperada do rio, está inferior a realidade, de acordo com a Resolução Conama. Além do mais, a região metropolitana de Curitiba, apresenta o IQA mais preocupante.

**Palavras-chave:** *Qualidade da Água, Estatística Multivariada, Paraná.*

---

<sup>1</sup>UFPR - e-mail: [gsartorik@gmail.com](mailto:gsartorik@gmail.com)

<sup>2</sup>UFPR - e-mail: [paulojus@ufpr.br](mailto:paulojus@ufpr.br)

## **Previsão da temperatura média para o município de Curitiba/PR utilizando Séries Temporais**

**Nicole Orsi<sup>1</sup>**

A temperatura influencia diversos processos como a agricultura e o comportamento dos animais, sendo assim, é de grande utilidade a sua previsão, permitindo a tomada de decisão com base nesta informação. Uma Série Temporal é o conjunto de observações feitas ao longo do tempo e, a partir da análise de suas componentes, como a sazonalidade, estacionariedade e tendência, é possível realizar a previsão do comportamento de dados. O presente trabalho tem como objetivo a previsão da temperatura média para o município de Curitiba/PR, para os anos de 2018 e 2019, a partir da análise e construção de Séries Temporais. Para a construção da Série Temporal antecedente foram utilizados dados dos anos de 2008 a 2017, obtidos do INMET. A estacionariedade foi testada pelo teste de KPSS, e a sazonalidade e tendência pelos testes de Friedman e de Dickey-Fuller Aumentado, respectivamente. A previsão foi executada utilizando Modelos Autorregressivos Integrados de Médias Móveis (ARIMA). De acordo com os testes a Série Temporal para a temperatura média possui estacionariedade e sazonalidade, e não possui tendência. Para 2018, a média da temperatura média prevista para o verão é de aproximadamente 21°C, variando entre 17°C e 25°C e, para o inverno, é de 15°C, variando entre 11°C e 20°C. Já para 2019, a média prevista para o verão é de aproximadamente 20°C, variando entre 16°C e 24°C e, para o inverno, a média é de 16°C, variando entre 12°C e 19°C. Pôde-se concluir que a Série Temporal analisada para Curitiba/PR apresenta um comportamento bastante sazonal.

**Palavras-chave:** *Clima, Séries Temporais, Temperatura Média.*

---

<sup>1</sup> Acadêmica de Agronomia – Universidade Federal de Santa Catarina (UFSC), Campus de Curitibanos. E-mail: nicoleorsi11@gmail.com.

## Recursos SAS para importação de grandes conjuntos de dados climáticos

Aline de Holanda Nunes Maia<sup>1</sup>

Ana Maria Heuminski de Ávila<sup>2</sup>

A análise de dados climáticos para previsões sazonais ou estudos de mudanças climáticas requer muitas vezes, a importação de dados disponíveis num grande número de arquivos, geralmente organizados por mês. Cada arquivo contém as informações diárias ou mensais de uma variável climática, total mensal de chuva ou temperatura média do mês, para diversas localizações geográficas na região de estudo. Neste trabalho apresentamos um programa no qual diversas funções SAS<sup>®</sup> são utilizadas para importação conjunta de múltiplos arquivos, sua concatenação e seleção de variáveis de interesse. O programa permite a extração eficiente de informações como nome da estação meteorológica e suas coordenadas, contidas no cabeçalho de cada arquivo individual e a importação os dados climáticos de cada estação, num determinado intervalo de tempo, para posterior concatenação. Como exemplo, apresentamos o caso de importação de dados mensais de chuva de vinte e nove locais (estações meteorológicas) no estado de São Paulo, para um estudo de caracterização espaço-temporal da estação chuvosa no estado e obtenção de previsões probabilísticas de início, duração e fim da estação.

**Palavras-chave:** *Dados climáticos, Pré-processamento de dados, funções SAS<sup>®</sup>*

---

<sup>1</sup> Embrapa Meio Ambiente, Jaguariúna, SP. e-mail [aline.maia@embrapa.br](mailto:aline.maia@embrapa.br)

<sup>2</sup> Cepagri – Unicamp, Campinas, SP. e-mail [avila@cpa.unicamp.br](mailto:avila@cpa.unicamp.br)

# REDES NEURAIS MLP VERSUS MODELO TARSO PARA PREVISÃO DE ENCHENTES NO RIO CUIABÁ

Lucas Monteiro Bianchi <sup>1 2 3</sup>

Anderson Castro Soares de Oliveira <sup>4</sup>

Marcelino Alves Rosa de Pascoa <sup>5</sup>

Luciana Sanches

Analia Macedo

A previsão da cota da bacia do rio Cuiabá é importante o desenvolvimento de métodos que possam auxiliar órgãos de defesa civil nas tomadas de decisões, resultando em maior segurança para a população, principalmente àquelas que residem próximo ao leito do rio. O conjunto de dados utilizados consistem em 4383 observações diárias das cotas médias (cm) e a precipitação (mm) da estação Rosário Oeste no período de janeiro de 2001 a dezembro de 2012. Foi feito um estudo comparativo utilizando e não utilizando os dados da precipitação para redes neurais perceptron multicamadas (MLP) e para o modelo TARSO com estimação bayesiana. Foram comparadas 78 redes neurais distintas, todas elas com 7 dias de defasagem no tempo visando prever um passo à frente. Para o modelo TARSO, foram consideradas 10000 iterações para cada cadeia sendo que as 2000 primeiras foram desprezadas e, das observações restantes, uma foi tomada a cada 10, de modo a obter uma amostra independente e identicamente distribuída. O retardo foi tomado igual a 1. Após a divisão feita nos dados pela variável threshold, o ajuste do modelo do primeiro regime foi feito com 2239 observações e o ajuste do segundo com 2144. Para efeito de comparação, os erros gerados pelas previsões foram mensurados através do MAPE, sendo selecionado como a melhor metodologia àquela que apresentar menor valor para o MAPE. Considerando ou não os dados de precipitação, as redes neurais MLP apresentaram menores erros de predição em relação ao modelo TARSO.

**Keywords:** *Redes neurais; Modelo TARSO, Séries temporais.*

---

<sup>1</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

<sup>2</sup>Agradecimento à FAPEMAT pela bolsa PIBIC durante a graduação.

<sup>3</sup>PPESTBIO - UFV. E-mail: *lucas\_bianchi123@hotmail.com*

<sup>4</sup>ICET - UFMT. E-mail: *andersoncso@ufmt.br*

<sup>5</sup>ICET - UFMT. E-mail: *marcelino.pascoa@gmail.com*

# Seleção dos melhores parâmetros do inverse distance weighting para estimar a precipitação média anual de Pernambuco, Brasil

Antonio Samuel Alves da Silva <sup>1</sup>

Borko Stosic <sup>1</sup>

Rômulo Simões Cezar Menenizes <sup>2</sup>

O inverse distance weighting (idw) é uma técnica robusta para interpolação de dados georreferenciados. Esta técnica pressupõe que cada ponto amostral tem influência sobre o ponto estimado e que esta influência diminui com o aumento da distância. O ajuste das estimativas no idw é controlado por um parâmetro  $p$  (quanto maior o valor de  $p$  menor a influência dos pontos mais distantes), geralmente escolhido de forma arbitrária (ou como  $p = 2$ ), levando a estimativas pouco confiáveis para variável em estudo. Além disso, uma outra característica relevante é o número de pontos vizinhos ( $n$ ) utilizados na estimativa. Com o intuito de identificar quais valores de  $p$  e  $n$  melhor representam a distribuição espacialmente explícita da precipitação de Pernambuco, foram utilizados dados (133 estações) de precipitação média anual (1950-2012) cedidos pelo Instituto de Tecnologia de Pernambuco (ITEP). O processo de validação cruzada permitiu fazer a comparação das interpolações geradas e as medidas *mean absolute error* ( $mae$ ), *root mean square error* ( $rmse$ ), *Nash-Sutcliffe efficiency* ( $nse$ ), *index of agreement* ( $d$ ), *coefficient of determination* ( $r^2$ ), foram utilizadas para avaliar/selecionar quais os melhores valores de  $p$  e  $n$  para interpolar a precipitação média anual. Os valores de  $p$  e  $n$  para os quais obteve-se os melhores (menores) valores de  $mae$  (96,17),  $rmse$  (149,89), e os melhores (maiores) valores de  $nse$  (0,72),  $d$  (0,91),  $r^2$  (0,72), foram respectivamente  $p = 1, 2$  e  $n = 5$ . O procedimento empregado para escolha de  $p$  e  $n$  mostrou-se como uma alternativa robusta para escolha desses parâmetros.

**Palavras-chave:** *Inverse distance weighting; cross validation; Precipitação, Pernambuco.*

---

<sup>1</sup>DEINFO - UFRPE. e-mail: [samueldmatematico@yahoo.com.br](mailto:samueldmatematico@yahoo.com.br)

<sup>1</sup>DEINFO-UFRPE

<sup>2</sup>DEN-UFPE.

## Uso de modelos logísticos na avaliação de impacto causado pelo desmanche de rocha em Lavras, MG

Vânia de Fátima Lemes de Miranda <sup>1 4</sup>

Izabela Regina C. de Oliveira <sup>2 4</sup>

Jefferson Medeiros Domingues <sup>3 4</sup>

Denise Caroline Rezende <sup>3 4</sup>

Yara Santos Lares <sup>3 4</sup>

Saulo Rocha Ferreira <sup>3 4</sup>

A mineração inclui processos e atividades industriais o objetivo principal é a extração de substâncias minerais de depósitos ou massas minerais. Como atividade industrial, a mineração é essencial para o avanço da sociedade. No entanto, a extração mineral, como qualquer atividade humana, afeta o meio ambiente e a vida das pessoas. Como impactos positivos, destacamos a empregabilidade e o fornecimento de matérias-primas para fins de infraestrutura. Por outro lado, os impactos negativos incluem: desconforto causado por tremores decorrentes de explosivos, impactos ambientais causados pela exploração de minérios, danos à saúde dos trabalhadores e danos aos edifícios ao redor das minas. A avaliação de potenciais efeitos e intensidade do impacto da extração mineral circundantes fundamental, pois ações podem ser tomadas para reduzir a insatisfação dos residentes. Neste trabalho, estudamos a percepção dos moradores em três bairros de Lavras, MG, sobre as atividades de uma mineradora situada nas proximidades. As residências foram estratificadas de acordo com a proximidade com a mineradora, resultando em três níveis de distância, chamados I (o mais próximo), II e III (o mais distante). Diferentes aspectos foram investigados e o efeito (fixo) da distância nas respostas dos residentes foram avaliados por modelos de regressão logística. Na maioria dos casos, percebemos diferenças significativas entre as áreas I e III, sendo essa última a que forneceu respostas mais favoráveis em relação à satisfação em relação à atividade mineradora.

**Palavras-chave:** *Modelos Lineares Generalizados, Mineração, Construção Civil.*

---

<sup>1</sup>DES - UFLA. e-mail: [vaniaflm@posgrad.ufla.br](mailto:vaniaflm@posgrad.ufla.br)

<sup>2</sup>DES - UFLA.

<sup>3</sup>DEG - UFLA.

<sup>4</sup>Agradecimento ao CNPq pelo apoio financeiro.

# Verifying ballast water compliance with international standards: a Bayesian approach

Eliardo Guimarães Costa <sup>1 4</sup>

Carlos Daniel Paulino <sup>2</sup>

Julio M. Singer <sup>3 4</sup>

We employ a Bayesian approach to compute sample sizes for estimating the organism concentration in ballast water. Both average coverage and average length of credible intervals under a Poisson (negative binomial) model with gamma (Pearson type VI) prior distributions are employed. We also consider a simulation study to evaluate the proposed methods and provide some practical guidelines. The Bayesian approach is more flexible than the corresponding frequentist approach since it naturally provides a sequential way of enhancing the estimation procedure along the deballasting process.

**Palavras-chave:** *Average coverage criterion, Average length criterion, Negative binomial distribution, Poisson distribution, Sample size.*

---

<sup>1</sup>Despartamento de Estatística - UFRN. e-mail: [eliardocosta@gmail.com](mailto:eliardocosta@gmail.com)

<sup>2</sup>Instituto Superior Técnico, Universidade de Lisboa e CEAUL-FCUL. e-mail: [cd.paulino@gmail.com](mailto:cd.paulino@gmail.com)

<sup>3</sup>Departamento de Estatística, USP. e-mail [jmsinger@ime.usp.br](mailto:jmsinger@ime.usp.br)

<sup>4</sup>This research received financial support from CNPq (3304126/2015-2) and FAPESP (2013/21728-2).

Outra área de aplicação



# Abordagens Clássica e Bayesiana para Polinômios Fracionários

Dennison Célio de Oliveira Carvalho <sup>1 2</sup>

Miriam Harumi Tsunemi <sup>2</sup>

Luzia Aparecida Trinca <sup>2</sup>

Os Polinômios Fracionários (PF) foram propostos em 1994 como uma extensão dos modelos de regressão polinomial, sob o ponto de vista da inferência clássica. O uso destes modelos é interessante quando a suposição de linearidade entre a variável resposta e as covariáveis não é válida. Em 2011 foi implementado o paradigma Bayesiano para os Polinômios Fracionários utilizando a priori *hiper-g* para garantir previsões consistentes, visto que, esta priori leva a uma forma fechada da verossimilhança marginal  $f(\mathcal{D}|\boldsymbol{\theta})$ . Ambas abordagens são alternativas para seleção de covariáveis de um modelo, bem como na designação da sua forma funcional. Neste trabalho, utilizaram-se alguns bancos de dados disponíveis na literatura com o objetivo de explorar estas técnicas através dos pacotes *mfp* e *bfp* disponíveis no *software* R. Diversos ajustes foram comparados a partir da raiz do erro quadrático médio de previsão (RMSPE) do modelo PF, da máxima densidade a posteriori (MAP) e da ponderação bayesiana de modelos (BMA), cujas previsões foram obtidas pela média de 3.000 previsões específicas do modelo.

**Palavras-chave:** *Priori hiper-g, MAP, BMA, RMSPE.*

---

<sup>1</sup>Agradecimento a CAPES pelo apoio financeiro. E-mail: dennison.carvalho@ibb.unesp.br

<sup>2</sup>Universidade Estadual Paulista (Unesp), Instituto de Biociências, Botucatu.

# Algoritmos para Predição de Processos Movidos Contra Grandes Litigantes

Jhenifer Caetano Veloso <sup>1</sup>

Eduardo Elias Ribeiro Junior <sup>2</sup>

Empresas envolvidas como réu em um alto número de ações são conhecidas como grandes litigantes, entre elas estão bancos e prestadoras de serviços de telefonia. Dado o atual grande número de processos em trâmite, a presente pesquisa visa auxiliar a estratégia utilizada pelo autor da ação, a fim de diminuir a quantidade e o tempo das ações. Para tanto, são propostos modelos preditivos, baseados na distribuição multinomial, para o resultado das ações. Os modelos são ajustados pelo método da máxima verossimilhança e a avaliação de características relevantes ao resultado da ação é feita pelo teste de razão de verossimilhanças. A base de dados que motiva esse trabalho tem origem na Consulta de Processos de Primeiro Grau do Tribunal de Justiça de São Paulo e é composta de sentenças, homologadas em 2014, em primeira instância de processos movidos contra grandes litigantes e é composta por 12.616 observações e 12 variáveis, sendo o resultado da ação a variável resposta. Os resultados do trabalho mostraram que empresa envolvida, foro, tipo de vara, tipo de dano, e os indicativos de presença de advogado do réu, discussões a respeito dos órgãos de proteção ao crédito, presença da palavra “terceiro”, discussões a respeito de relação de consumo e presença de discussões a respeito de gratuidade judiciária são características relevantes para o resultado da ação. O modelo obteve uma probabilidade de acerto de 53% na base de teste (25% dos dados), indicando uma capacidade preditiva concordante com outros modelos também ajustados a mesma base de dados.

**Palavras-chave:** *Jurimetria, Distribuição Multinomial, Predição.*

---

<sup>1</sup>Programa de Educação Tutorial - Estatística - UFPR. e-mail: [jhenicaet@gmail.com](mailto:jhenicaet@gmail.com)

<sup>2</sup>Departamento de Ciências Exatas (LCE) - ESALQ-USP

## **Amostragem sequencial na detecção preferência por refrigerantes sabor guaraná entre consumidores de Alfenas, MG**

**Amanda Felix Camocardi<sup>1 5</sup>**

**Eric Batista Ferreira<sup>2 5</sup>**

**Cássia de Souza Santos<sup>3 5</sup>**

**Gabriel Alex da Silva Alves<sup>4 5</sup>**

A amostragem sequencial é uma técnica caracterizada pela tomada de sucessivos elementos da população que permitem a parada no tamanho de amostra mínimo suficiente para se tomar uma decisão com taxas de erro previamente estabelecidas. No Brasil, o início do processo de industrialização de bebidas à base de guaraná ocorreu a partir de 1905 e, de acordo com dados das pesquisas realizadas pelo IBGE, seu consumo não para de crescer. Portanto, o objetivo deste estudo é analisar, no município de Alfenas, MG, a preferência por refrigerantes do tipo guaraná das marcas Antártica e Kuat, a partir de estratificação socioeconômica. Três supermercados, frequentados predominantemente por consumidores de alta (>5SM), média (entre 5 e 3SM) e baixa (<3SM), foram visitados. Utilizando-se da amostragem sequencial e respeitando os níveis de erro tipo I e tipo II preestabelecidos a 5%, no supermercado de classe alta, observou-se que com apenas 26 participantes já foi possível determinar que não há preferência entre marcas, isto é, pelo teste cego elas não são identificáveis; no de classe média com 29 participantes tomou-se a decisão de não preferência; e no de classe baixa com 34 participantes, o mesmo resultado. Sendo assim, notou-se o mesmo comportamento, independente de classe socioeconômica. Embora os consumidores afirmem preferir a marca Antártica pelo sabor, não foram capazes de distingui-la da marca Kuat em testes cegos. Isso sugere que outros aspectos, como marketing e preço podem estar influenciando a preferência.

**Palavras-chave:** *Análise sensorial, Segmentação de mercado, Testes de hipóteses.*

---

<sup>1</sup> Graduanda em Nutrição, UNIFAL-MG. e-mail amandacamocardi@gmail.com

<sup>2</sup> Professor do Departamento de Estatística, Instituto de Ciências Exatas, UNIFAL-MG.

<sup>3</sup> Mestranda em Estatística Aplicada e Biometria, UNIFAL-MG.

<sup>4</sup> Bolsistas do Programa de Iniciação Científica Júnior, Ensino Médio.

<sup>5</sup> Agradecimento ao CNPq, FAPEMIG e Capes, pelo apoio financeiro.

---

## **Análise da série de violência contra a mulher em São João del-Rei**

**Denise de Assis Paiva<sup>1</sup>**

**Rejane Corrêa da Rocha<sup>2</sup>**

**Luciane Teixeira Passos Giarola<sup>2</sup>**

Esse trabalho teve como objetivo analisar o comportamento da série de violência contra a mulher, em São João del-Rei, por meio da metodologia de séries temporais. Os dados de criminalidade foram coletados junto à Delegacia Adjunta de Repressão a Crimes contra a Mulher, da Polícia Civil de Minas Gerais (PCMG), regional de São João del-Rei, sendo considerada a soma mensal dos crimes classificados no Registro de Eventos da Defesa Social (REDS) que são: assédio sexual, estupro, estupro de vulnerável, outras infrações contra a dignidade sexual e da família. A série é composta de 47 observações, correspondendo ao período de janeiro de 2014 a novembro de 2017. Foi utilizado o software Gretl e o único modelo ajustado aos dados foi o de média móveis degenerado MA[(10)]. Concluiu-se que a metodologia de séries temporais é útil para descrever a série de violência contra a mulher em São João del-Rei.

**Palavras-chave:** *Crimes contra mulher, modelos de Box & Jenkins, séries temporais.*

---

<sup>1</sup> DES – UFLA. denisekpela@hotmail.com

<sup>2</sup> DEMAT – UFSJ. rejane@ufs.edu.br

<sup>3</sup> DEMAT – UFSJ. luciane@ufs.edu.br

## Análise de dados na forma de contagens longitudinais com superdispersão na produção de morango

Sidleide Barbosa de Sousa <sup>1 3</sup>

Clarice Garcia Borges Demétrio <sup>2 3</sup>

Giovana Fumes <sup>3 3</sup>

Fernanda Canassa <sup>4 3</sup>

Italo Delalibera Junior <sup>5 3</sup>

Na experimentação agronômica, dados na forma de contagens são frequentemente obtidos ao longo do tempo. Para analisar dados com tais características, a distribuição de Poisson pertencente à família exponencial é amplamente utilizada. Entretanto, dados longitudinais na forma de contagens podem apresentar uma variabilidade maior do que a captada por tal modelo, fenômeno chamado de superdispersão. Tal característica em dados longitudinais pode acontecer devido à relação entre média e variância ou ainda devido à correlação existente entre as medidas repetidas na mesma unidade experimental. Desse modo, ignorar a presença de superdispersão nos dados pode conduzir a uma interpretação errada dos resultados. Molenberghs et al. (Statistical Science, 25:325-347, 2010) propuseram modelos lineares generalizados mistos estendidos (MLGME) a fim de modelar a superdispersão e a correlação. Este trabalho tem por objetivo estudar dados longitudinais na forma de contagens com superdispersão, usando um conjunto de dados de um experimento casualizado em blocos, o qual busca avaliar os efeitos benéficos às plantas de morango inoculadas com diferentes fungos, estudando diversas variáveis, como números de folhas e flores no período de 0, 30, 60, 90, 120, 150 e 180 dias após inoculados. Tais modelos acomodam a superdispersão e a correlação entre medidas obtidas ao longo do tempo.

**Palavras-chave:** *Distribuição de Poisson; Fungos entomopatogênicos; Modelos lineares generalizados mistos estendidos.*

---

<sup>1</sup>ESALQ - USP. e-mail: sbarbosas1987@usp.br

<sup>2</sup>ESALQ - USP. e-mail: clarice.demetrio@usp.br

<sup>3</sup>ESALQ - USP. e-mail: gifumesbtu@gmail.com

<sup>4</sup>ESALQ - USP. e-mail: fernanda.canassa@usp.br

<sup>5</sup>ESALQ - USP. e-mail: delalibera@usp.br

<sup>3</sup>Agradecimento a CAPEs pelo apoio financeiro.

## **Análise de potencial eólico em Fernando de Noronha-PE utilizando distribuições de probabilidade**

**Fábio Sandro dos Santos<sup>1</sup>**

**Kerolly Kedma Felix do Nascimento<sup>23</sup>**

**Diego Alves Gomes<sup>23</sup>**

**Jader da Silva Jale<sup>23</sup>**

**Tiago Alessandro Espínola Ferreira<sup>2</sup>**

Na busca por alternativas viáveis para preencher a crescente demanda por energia, encontra-se a energia eólica como uma alternativa limpa para produção e geração da mesma, uma vez que este recurso é inesgotável. Neste trabalho, analisou-se a velocidade média horária do vento na estação meteorológica de Fernando de Noronha-PE (INMET) no Nordeste brasileiro, entre os anos de 2004 a 2017, a uma altura de 10m. Com o propósito de verificar o potencial de energia eólica, foram utilizadas as seguintes Funções Densidade de Probabilidade (FDP's): Weibull-2p, log-Normal, Gamma e Burr. Os parâmetros de forma ( $k$ ) e de escala ( $c$ ) das distribuições aqui citadas são determinados com base na velocidade média diária medida na série de velocidade. Foram utilizados os critérios de seleção AIC e BIC e as estatísticas MAPE, MAD, MSD e RMSE, a fim de comparar os ajustes das diferentes FDP's e verificar a que melhor se ajustou aos dados de velocidade do vento. Os resultados indicam que as distribuições Weibul-2p e a Burr foram as que melhor se ajustaram aos dados analisados. Além disso, observou-se que a Região de Fernando de Noronha, nos meses de agosto a novembro, apresentou uma alta incidência de vento, com uma frequência acima de 4m/s. Assim sendo, a região apresenta recursos eólicos satisfatórios para a instalação de parque eólico.

**Palavras-chave:** *Ajuste, distribuições, velocidade do vento.*

---

<sup>1</sup>PPGBEA – UFRPE. e-mail: fabio.sandropb@gmail.com

<sup>2</sup>PPGBEA – UFRPE.

<sup>3</sup>Agradecimento ao Capes pelo apoio financeiro.

# Análise de precipitação em Campina Grande-PB por cadeia de Markov de ordem superior

Kerolly Kedma Felix do Nascimento <sup>1 3</sup>

Fabio Sandro dos Santos <sup>2</sup>

Rosendo Chagas de Albuquerque <sup>2 3</sup>

Jader da Silva Jale <sup>2 3</sup>

Tiago Alessandro Espínola Ferreira <sup>2</sup>

Neste trabalho, analisamos dados de precipitação diária de uma estação meteorológica localizada na cidade de Campina Grande-PB, no período compreendido entre 02/02/1963 e 13/12/2017, totalizando 14913 dias. Utilizamos modelos de cadeia de Markov, a tempo discreto, com dependência de até sexta ordem ( $r = 7$ ), com dois estados possíveis ( $m = 2$ ): dia seco (precipitação menor que  $1,0mm$ ) e dia chuvoso (precipitação de pelo menos  $1,0mm$ ). Utilizamos os critérios de informação de Akaike (AIC), Akaike corrigido (AICc), Bayesiano (BIC) e Hannan-Quinn (HQC) para estimar a ordem ótima  $k^*$  dentre os modelos de cadeia de Markov, por meio da maximização da razão de verossimilhanças entre esses modelos, em que  $k < r$ , com  $k = 0, 1, \dots, r - 1$ . Nossos resultados indicam que a ordem ótima encontrada é  $k^* = 5$ . Isso quer dizer que, com base no limiar de chuva de  $1,0mm$ , a probabilidade de um dia qualquer ser seco ou chuvoso depende apenas das condições climáticas observadas nos últimos 5 dias anteriores. Além disso, estimamos a matriz de transição de probabilidades condicionais com dimensões  $m^{k^*} \times m = 2^5 \times 2$ .

**Palavras-chave:** *Cadeia de Markov, Campina Grande, ordem superior, precipitação.*

---

<sup>1</sup>DEINFO - UFRPE. e-mail: [kerollyfn@gmail.com](mailto:kerollyfn@gmail.com)

<sup>2</sup>DEINFO - UFRPE.

<sup>3</sup>Agradecimento à CAPES pelo apoio financeiro.

## Análise do Desempenho no Ensino Médio na Área Metropolitana de Brasília: Uma Abordagem Multinível

Davi Souza Botelho <sup>1</sup>

Ana Maria Nogales Vasconcelos <sup>2</sup>

Maria Teresa Leão Costa <sup>3</sup>

Este estudo teve como objetivo verificar em que medida as características dos estudantes e da escola podem influenciar no desempenho dos estudantes do ensino médio, participantes do Exame Nacional do Ensino Médio em 2014, de escolas localizadas no Distrito Federal ou em sua periferia metropolitana. Os dados utilizados foram extraídos do ENEM e Censo Escolar de 2014 realizados pelo INEP. Foram selecionados os estudantes concluintes do ensino médio público regular e que residiam e estudavam na Área Metropolitana de Brasília (AMB). Para averiguar os fatores associados ao desempenho dos alunos utilizaram-se os Modelos Multiníveis, onde os estudantes constituíam o primeiro nível e as escolas o segundo. Os resultados no modelo nulo mostram uma média geral no desempenho dos alunos de 494,5 e um coeficiente de correlação intraclasses igual a 0,18, que indica, a princípio, que 18% da variação do desempenho dos alunos está associado à variabilidade entre as escolas. O modelo final resultou que os atributos que interferem no desempenho são: o sexo, a idade, a cor, a escolaridade dos pais e a defasagem para o primeiro nível e para o segundo nível: o nível socioeconômico, taxa de participação no ENEM, o indicador de formação docente e média de alunos por turma. Ainda que sejam preliminares, os resultados aqui apresentados evidenciam o potencial dos dados produzidos pelo INEP/MEC para a avaliação educacional no país.

**Palavras-Chave:** *Fatores Associados ao Desempenho; Exame Nacional do Ensino Médio; Teoria de Resposta ao Item; Modelos Lineares Multiníveis; Indicadores Educacionais; Área Metropolitana de Brasília.*

---

<sup>1</sup>Bacharel em Estatística pela UnB e mestrando na ENCE/IBGE e-mail: [daviestatistico@gmail.com](mailto:daviestatistico@gmail.com)

<sup>2</sup>Professora do Departamento de Estatística da UnB. e-mail: [nogales@unb.br](mailto:nogales@unb.br)

<sup>3</sup>Professora do Departamento de Estatística da UnB. e-mail: [mtleao@unb.br](mailto:mtleao@unb.br)



# Análise Espacial e Multivariada da Esperança de Vida ao Nascer

Walef Machado de Mendonça <sup>1 3</sup>

Leonardo Biazoli <sup>2 3</sup>

Ana Carolina de Oliveira <sup>2 3</sup>

Este trabalho busca analisar a distribuição espacial da esperança de vida ao nascer (EVN) nos municípios de Minas Gerais e, identificar se as localidades onde a EVN é superior são aquelas com maior desenvolvimento socioeconômico. Os dados utilizados são provenientes do Censo Demográfico 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e consultados a partir do Índice Mineiro de Responsabilidade Social (IMRS). Para a análise da distribuição espacial da esperança de vida foi empregada a análise exploratória de dados espaciais (AEDE) para descrever os padrões de associação espacial através do coeficiente de autocorrelação I de Moran. As regiões com maior desenvolvimento socioeconômico foram identificadas por meio da análise de agrupamento. Aplicou-se o método de Ward para definição do número de grupos e, visualmente, através do ponto de corte de maior distância do dendrograma, definiu-se esse número como 4. Posteriormente, o método não hierárquico das k-médias alocou os municípios em 4 grupos. Observa-se que o Grupo 4 apresenta uma menor variabilidade relativa dos dados o que aponta para uma semelhança no comportamento das variáveis nesses municípios. Como resultado a AEDE aponta que há a existência de dependência espacial, ou seja, percebe-se que, os municípios com maior EVN tendem a estar próximos de outros municípios com essa mesma característica. Concomitantemente, através do agrupamento, o Grupo 4, considerado o mais desenvolvido socioeconomicamente, engloba grande parte dos municípios pertencentes ao cluster obtido pela AEDE, que reúne as localidades com as maiores EVN.

**Palavras-chave:** *Análise exploratória de dados espaciais, Agrupamento, Esperança de vida ao nascer.*

---

<sup>1</sup>Universidade Federal de Alfenas - UNIFAL-MG. e-mail: *walefm2@gmail.com*

<sup>2</sup>Universidade Federal de Alfenas - UNIFAL-MG.

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

---

**Análise Multivariada de Perfis de manejos de irrigação e doses de nitrogênio na cultura do arroz (*Oryza sativa*)****Vinícius Henrique Beltran<sup>1</sup>****André Fróes de Borja Reis<sup>2</sup>****José Laércio Favarin<sup>3</sup>****Carlos Tadeu dos Santos Dias<sup>4</sup>**

A Análise Multivariada de Perfis (AMP) é uma excelente forma para analisar dados mensurados no tempo, mas também no espaço, tendo como objetivo testar hipóteses (paralelismo, coincidência e horizontalidade) sobre os perfis médios de resposta dos diversos tratamentos e compará-los entre si. No caso deste projeto, a metodologia da Análise Multivariada de Perfis foi utilizada para analisar um conjunto de dados de medidas repetidas ao longo do tempo proveniente de um experimento com cultura de arroz (*Oryza sativa*), constituído por duas safras (2014/2015 e 2015/2016), as quais foram submetidas a cinco manejos de irrigação (aplicados em parcelas) e dois níveis de nitrogênio (aplicados em subparcelas). Para este experimento, foram avaliadas quatro variáveis ao longo do tempo: Matéria Seca da Parte Aérea, Acúmulo de Nitrogênio, Teor de Nitrogênio e Índice de Área Foliar.

A partir dos resultados obtidos, foi observado que, em ambas as safras e para as quatro variáveis em estudo, as hipóteses de paralelismo e coincidência não foram rejeitadas, de modo que os perfis são paralelos e coincidentes entre si, enquanto a hipótese de horizontalidade foi rejeitada, indicando que os mesmo perfis médios em todas estas condições não foram horizontais. Além disso, também foram construídos contrastes para comparar os perfis médios dois a dois, os quais mostraram que o perfil médio SL 150 (manejo de irrigação: sem irrigação suplementar – SL; dose de nitrogênio: 150 kg.ha<sup>-1</sup>) foi superior em valores do perfil médio a todos os demais, em ambas as safras e para todas as variáveis estudadas.

**Palavras-chave:** Análise Multivariada de Perfis; dados de medidas repetidas; *Oryza sativa*.

---

<sup>1</sup> LCE – ESALQ/USP. E-mail: [vinicius.beltran@usp.br](mailto:vinicius.beltran@usp.br)

<sup>2</sup> LPV – ESALQ/USP. E-mail: [andrefbr@usp.br](mailto:andrefbr@usp.br)

<sup>3</sup> LPV – ESALQ/USP. E-mail: [favarin.esalq@usp.br](mailto:favarin.esalq@usp.br)

<sup>4</sup> LCE – ESALQ/USP. E-mail: [ctsdias@usp.br](mailto:ctsdias@usp.br)

# Aplicação da Análise de Cluster em Dados de Crimes nos EUA

Jessica Cristina Pereira Batista <sup>1</sup>

Ednário Barbosa de Mendonça<sup>1</sup>

Tiago Almeida de Oliveira <sup>1</sup>

Ricardo Alves de Olinda <sup>1</sup>

Marcelino Alves Rosa de Pascoa <sup>2</sup>

A análise de cluster é uma técnica estatística usada para classificar elementos em grupos de uma forma em que os elementos dentro de um mesmo cluster sejam muito parecidos, e elementos em clusters diferentes sejam distintos entre si. Sendo assim, os métodos hierárquicos da análise de cluster tem como principal característica a obtenção de um algoritmo que fornece a partição de um conjunto de dados em vários agrupamentos possíveis, onde um cluster pode ser mesclado a outro em determinado passo desse algoritmo. Neste trabalho, foi feita a análise de cluster do conjunto de dados USArrests, disponível no CRAN do software R pelo pacote *datasets*, o qual contém as estatísticas de prisões (por 100.000 habitantes) ocasionadas por assalto, assassinato e estupro, em cada um dos 50 estados dos EUA no ano de 1973. O objetivo do trabalho em questão consistiu em formar grupos a partir dos dados de crimes por meio da análise de cluster, utilizando o método hierárquico de seleção de grupos. Inicialmente, por meio do método de particionamento *lk-means*, foi obtido o número ótimo de clusters a serem considerados a partir da base de dados. Em seguida, foi feito o dendrograma dos clusters a serem formados, em que foi utilizada a distância euclidiana para a formação da matriz de distâncias e, para construir os ramos do dendrograma, foi utilizado o método de ligação direta (*complete-linkage*). Tendo-se formado os clusters, foi feita uma análise exploratória de cada cluster individualmente, obtendo uma descrição das características de cada cluster formado.

**Palavras-chave:** *análise de cluster; método hierárquico; distância euclidiana.*

---

<sup>1</sup>Departamento de Estatística - UEPB. e-mail:eddybarbosa92@gmail.com

<sup>2</sup>Departamento de Estatística - UFMT

# Assessing Robustness of Linear Model using Simulation

Danilo Augusto Sarti <sup>1</sup>

Carlos Tadeu dos Santos Dias <sup>2</sup>

This paper uses simulation in software to assess the robustness of linear models estimation. In paper we evaluate simulated data with linear regression, robust regression and quantile regression. We evaluate different scenarios considering outliers data and explore the fitted models and robustness consequences in big data, datascience and machine learning contexts. The results shows how robust regressions can be a valid alternative for models and predictions in the context of data-sets of interest.

**Keywords:** *Robustness, Simulation, Estimation.*

---

<sup>1</sup>LCE - USP/ESALQ. e-mail:danilo.estat@gmail.com

<sup>2</sup>LCE - USP/ESALQ. e-mail:ctsdias@usp.br

## ASSOCIAÇÃO ENTRE CARACTERÍSTICAS DE MULHERES VÍTIMAS E DOS AUTORES DE VIOLÊNCIA DOMÉSTICA DO MUNICÍPIO BELÉM-PA, BRASIL

Thays Suelen Brito Santos<sup>1 3</sup>

Edson Marcos Leal Soares Ramos<sup>1 3</sup>

Gelilza Salazar Costa<sup>2</sup>

Silvia dos Santos de Almeida<sup>1 3</sup>

**Resumo:** A violência doméstica é a violência ou outro tipo de abuso por parte de uma pessoa contra outra num contexto doméstico, como no caso de um casamento ou união estável, ou contra crianças e/ou idosos. Esse tipo de violência é considerado estritamente de cunho privado, pertencente ao lar, e aos poucos vem ganhando notoriedade na sociedade e tornando-se cada vez mais um assunto de caráter público e atual. O objetivo deste trabalho é de apresentar um estudo sobre as possíveis associações entre características de mulheres vítimas e dos autores de violência doméstica no município de Belém. Para tal, utilizam-se dados referentes aos crimes de violência doméstica sofrida por mulheres, registrados no Sistema Integrado de Segurança Pública do estado do Pará, no período de 2013 a 2015. Sendo aplicada a técnica estatística multivariada de Análise de Correspondência objetivando verificar as possíveis associações entre as categorias das características de mulheres vítimas e dos autores de violência doméstica. Dentre os principais resultados, pode-se destacar que as mulheres vítimas da violência doméstica, cujo estado civil é “casada”, apresentam associação estatística, em relação ao tipo de crime sofrido, com as categorias “sofre ameaça”, “perturbação da tranquilidade” e “violação de domicílio”. Acredita-se que a discussão acadêmica acerca da violência doméstica torna-se fundamental para compreender a dinâmica que a norteia e dá visibilidade ao tema, bem como contribuir para que as instituições públicas e toda a sociedade possam embasar e adequar políticas de prevenção, enfrentamento e combate a esse tipo de violência.

**Palavra-chave:** Crime Sofrido; Políticas de prevenção, Análise de Correspondência.

---

<sup>1</sup> ICEN – UFPA.

<sup>2</sup> FAEST – UFPA. gelilza@gmail.com

<sup>3</sup> Agradecimento ao CNPq pelo apoio financeiro.

## Avaliação da dieta do Pirarucu (*Arapaima gigas*) via Componentes Principais

Pedro Marinho Amoedo<sup>1</sup>

Rick Anderson Freire Mangueira<sup>2</sup>

Marcello Neiva de Mello<sup>2</sup>

Djair Durand Ramalho Frade<sup>2</sup>

Carlos Tadeu dos Santos Dias<sup>2</sup>

Com a finalidade de estudar o efeito do farelo de soja como alimento substituto parcial da farinha de peixe para espécie carnívora Pirarucu (*Arapaima gigas*), foi instalado um experimento com 105 juvenis da espécie em Manaus AM. Estes foram confinados em tanques e tratados por 120 dias com uma dieta formulada por cinco níveis de substituição (0, 15, 30, 45 e 60%) da farinha de peixe. Nove variáveis zootécnicas foram medidas para identificar o possível efeito da dieta. O objetivo do trabalho foi avaliar o efeito de níveis crescentes de substituição da proteína de farinha de peixe pela proteína do farelo de soja em dietas para juvenis de pirarucu. Para a análise estatística do experimento, utilizou-se Componentes Principais, Regressão Polinomial e Análise de Variância. Como resultados alcançados, foi possível constatar diferença significativa entre os níveis de substituição pelo teste de Wilks. Pôde-se verificar que níveis crescentes de substituição da farinha de peixe pelo farelo de soja proporcionou menor desempenho corporal em Pirarucu.

**Palavras-chave:** *Componentes principais, Dieta, Pirarucu.*

---

<sup>1</sup> LCE – ESALQ/USP. e-mail pamoedo@ufam.edu.br

<sup>2</sup> LCE – ESALQ/USP.

## **Avaliação do efeito do Própolis na cor de Silicones Protéticos Faciais Usando Análise de Variância Multivariada Não Paramétrica**

**Nathalia Bahia Miranda<sup>1</sup>**

**José Alcides Almeida de Arruda<sup>1</sup>**

**Sandra Barbosa Almeida<sup>2</sup>**

**Emerson Gomes dos Santos<sup>3</sup>**

**Igor Studart Medeiros<sup>2</sup>**

**Amália Moreno<sup>1</sup>**

As soluções brasileiras de extratos de própolis verde podem ser alternativas efetivas para desinfecção e controle da saúde de usuários de próteses maxilofaciais. No entanto, o efeito da própolis sobre a estética e as propriedades mecânicas dos elastômeros maxilofaciais precisa ser melhor investigado. O objetivo deste estudo in vitro foi avaliar o efeito da desinfecção com soluções de extrato de própolis verde brasileiras na alteração de cor dos valores transmitância, parâmetro de translucidez, relação de contraste de dois elastômeros maxilofaciais. Foram avaliados dois tipos de elastômeros maxilofaciais em 50 espécimes redondos (3x10 mm) de cada elastômero. Os espécimes foram submetidos a 60 dias de desinfecção, 3 vezes por semana. As soluções utilizadas foram: água destilada, clorexidina a 2%, e extratos de própolis verde a 11% (aquoso-PAQ, glicólico-PGL e alcoólico-PAL) por imersão durante 15 minutos. Os dados foram analisados com análise de variância multivariada não paramétrica (NPMANOVA), e testes de comparações múltiplas ( $\alpha=0,05$ ). Pode-se observar que os elastômeros desinfetados com soluções PAQ e PGL apresentaram valores de alteração de cor aceitáveis clinicamente. Pode-se verificar diferenças de opacidade e translucides para PAL e PGL no elastômero RTV. Conclui-se que soluções de própolis aquoso parece ser favorável para desinfecção sem comprometer as propriedades dos elastômeros. O elastômero HCR apresentou menor alteração das propriedades de cor comparado ao RTV.

**Palavras-chave:** *Prótese maxilofacial, Desinfecção, Própolis, Clorexidina, Cor, NPMANOVA.*

---

<sup>1</sup> Departamento de Clínica, Patologia e Cirurgia Odontológicas – UFMG

<sup>2</sup> Departamento de Biomateriais e Biologia Oral – USP

<sup>3</sup> Departamento de Administração – UNIFESP. Email: emerson.gomes@unifesp.br

## Avanços para o delineamento de experimentos em múltiplos estratos

Luzia Aparecida Trinca <sup>1</sup>

Experimentos em múltiplos estratos são frequentemente utilizados em diversas áreas do conhecimento como, por exemplo, biologia, agronomia e engenharia. Para reduzir os custos experimentais, em geral, os delineamentos não são ortogonais e portanto o desenvolvimento de métodos de construção de delineamentos eficientes é importante. A alternativa de construção sequencial, de estrato por estrato é bastante flexível, pode ser usada com qualquer critério e é aplicável à diversas estruturas. Neste trabalho exploramos as propriedades dos delineamentos usando critérios modificados que colocam peso na existência de graus de liberdade para erro puro em cada estrato. Aplicações nas áreas de bio processos e outras são utilizadas para ilustrar a flexibilidade do método.

**Palavras-chave:** *Delineamento ótimo, Split-plot, Strip-plot.*

---

<sup>1</sup>Bioestatística - IBB - Unesp. e-mail: [lutrinca@gmail.com](mailto:lutrinca@gmail.com)



## Bayesian inference for bivariate counts in a $2 \times 3$ completely random design: preliminary results

Everton Batista da Rocha <sup>1</sup>

Davi Butturi-Gomes <sup>2</sup>

Let  $X_i \sim \text{Poisson}(\lambda_i)$ ;  $i = \{0, 1, 2\}$ . Then, the random variables  $Y_1 = X_0 + X_1$  and  $Y_2 = X_0 + X_2$  jointly follow a bivariate Poisson (BiP) distribution with parameters  $\lambda_0$ ,  $\lambda_1$  and  $\lambda_2$ . Although such result is widely accepted, there are few studies using the BiP distribution in an experimental context. Thus, the objective of our study was evaluating the potential use of the BiP for a  $2 \times 3$  CRD with two count responses: the number of stomata on the upper and on the lower leaf surfaces. We used Bayesian methods for parameter estimation and inference, and due to the slow convergence of the chains, we considered only the main effects of the treatments in all of the three predictors. Logarithmic link functions were used in all cases as well. Our results showed that in one of the treatments ( $A_1B_0$ ) the correlation between the counts was strong enough to be taken into account whilst fitting the model for the response variables in the experiment.

**Palavras-chave:** *Bivariate Poisson distribution, Correlation modeling, MCMC.*

---

<sup>1</sup>IME - UFG. e-mail: [evertonbatista@ufg.br](mailto:evertonbatista@ufg.br)

<sup>2</sup>DEMAT - UFSJ. e-mail: [davibg@ufs.edu.br](mailto:davibg@ufs.edu.br)

## Caracterização dos municípios da Região Administrativa de Campinas por meio de técnicas multivariadas

Hercílio da Silva Freitas Júnior<sup>1 5</sup>

Welinton Yoshio Hirai<sup>2 5</sup>

Vivian Aparecida Brancaglioni<sup>3 5</sup>

Carlos Tadeu dos Santos Dias<sup>4 5</sup>

No Brasil existem formas distintas de se agrupar sua (ou determinada) área geográfica, utilizando características e critérios distintos. Uma dessas formas de agrupamento ocorre por meio de Regiões Administrativas, determinado pela similaridade econômica e social. O presente trabalho busca aglomerar e caracterizar os municípios da Região Administrativa de Campinas de acordo com variáveis socioeconômicas. O conjunto de dados utilizado é referente ao ano de 2016, proveniente da base de dados do IBGE. Os métodos utilizados foram as técnicas multivariadas: Análise de Cluster (AC) e Análise de Componentes Principais (ACP). Foram obtidos três aglomerados distintos de municípios utilizando a AC não hierárquica. Posteriormente as observações das aglutinações foram caracterizadas por meio da ACP, que utilizou três componentes, que juntos explicaram cerca de 76% da variância total dos dados.

**Palavras-chave:** *Agrupamento, Componentes principais, Municípios.*

---

<sup>1</sup>LCE - ESALQ/USP. e-mail: [hercilio.s.junior@usp.br](mailto:hercilio.s.junior@usp.br)

<sup>2</sup>LCE - ESALQ/USP. e-mail: [wyhirai@usp.br](mailto:wyhirai@usp.br)

<sup>3</sup>LCE - ESALQ/USP. e-mail: [vivian.brancaglioni@usp.br](mailto:vivian.brancaglioni@usp.br)

<sup>4</sup>LCE - ESALQ/USP. e-mail: [ctsdias@usp.br](mailto:ctsdias@usp.br)

<sup>5</sup>Agradecimento à CAPES pelo apoio financeiro.

## Células de Voronoi: Visualizando a distribuição de votos à um nível intramunicipal

Augusto Felix Marcolin <sup>1</sup>

Douglas R. Mesquita Azevedo <sup>1</sup>

Lucas da Cunha Godoy <sup>2</sup>

Luís Gustavo Silva e Silva <sup>1</sup>

Os resultados eleitorais são de grande importância tanto para partidos políticos quanto para a população em geral, tais dados estão disponíveis para todos em diversos níveis de agregação, sendo o menor deles a seção eleitoral. O acesso à quantidade de votos por seção eleitoral por sua vez permite a integração de dados socioeconômicos dos setores censitários trazendo informações relevantes sobre o perfil dos eleitores. Neste trabalho utilizou-se a técnica da tesselação de Voronoi para a criação de regiões no entorno das seções eleitorais com o objetivo de caracterizar o resultado das eleições em um nível mais granular. Os resultados desta aplicação são apresentados através de mapas interativos no qual o usuário pode realizar a investigação do resultado das eleições dentro de um município.

**Palavras-chave:** *Tesselação de Voronoi, Eleições, Visualização de dados.*

---

<sup>1</sup>DEST - UFMG.

<sup>2</sup>DEST - UFMG. e-mail: [lucasdac.godoy@gmail.com](mailto:lucasdac.godoy@gmail.com)

## Comparação da probabilidade de cobertura entre procedimentos de estimação para diferenças entre medias via simulação monte Carlo.

Rodnei Alves Marques<sup>1</sup>  
Vânia de Fátima Lemes de Miranda<sup>2</sup>  
Daniel Furtado Ferreira<sup>3</sup>

Uma das dificuldades comumente encontradas em pesquisas estatísticas é a violação da suposição de normalidade dos dados ou resíduos. Em muitos casos, as distribuições dos dados de duas amostras são não normais e, ainda, exibem falta de homogeneidade de variâncias. O intervalo de confiança é um dos conceitos mais utilizado e bem-sucedido na inferência estatística. O presente estudo tem como objetivo comparar a probabilidade de cobertura e amplitude médias dos intervalos de confiança obtidos a partir da estatística dos testes *t*-Student, Welch, Hall com médias aparadas e Hall sem médias aparadas para a diferença de duas médias de populações normais e não normais. Para isso, foram geradas amostras obtidas de populações normais, exponenciais e qui-quadrado via simulação Monte Carlo, considerando os seguintes parâmetros sem perda de generalidade: variância da população 1 igual a 1 e variância da população 2 igual a  $\delta$  vezes a variância população 1, com  $\delta$  variando da seguinte forma: 1, 2, 4, 8 e 16. Para cada uma dessas situações, as populações foram geradas com médias iguais e com médias diferentes, sendo as diferenças padronizadas entre as médias populacionais em 1, 2, 3 e 4 erros padrões. Foram obtidos 1.000 intervalos de confiança para cada combinação citada e, assim, obtido a probabilidade de cobertura e amplitude média, em que se constatou um bom desempenho dos intervalos de confiança obtidos a partir da distribuição *t*-Student e Hall com médias aparadas.

**Palavras chave:** Intervalos de confiança, Probabilidade de Cobertura, Simulação Monte Carlo.

1- Professor do Instituto Federal de Minas Gerais e aluno da pós graduação UFLA

2- Professora da Universidade Federal de Uberlândia e aluna da pós graduação UFLA

3- Professor da Universidade Federal de Lavras

## Comparação de modelos epidemiológicos para propagação de informações

Michele Martins Lopes <sup>1 4</sup>

Luiz Henrique Marra da Silva Ribeiro <sup>2 4</sup>

José Paulo Carvalho dos Santos <sup>3 4</sup>

O estudo da propagação de informações hoje em dia é muito útil. Diversos modelos para tal propagação são propostos com base no modelo SIR da epidemiologia. Este trabalho consta de uma comparação entre dois modelos epidemiológicos propostos por dois autores diferentes. Foram feitas quatro simulações no software R para realizar a comparação, em relação à quatro diferentes assuntos, sendo que cada um apresentou uma diferente forma de propagação nos períodos em que foram analisados. Para as simulações foi utilizada a ferramenta “Google Trends” para se obter os dados já observados, a fim de compará-los com os obtidos pelos dois modelos analisados. Logo após, foram calculados dois tipos de erro em cada simulação, para ambos os modelos. A partir disso, pôde-se concluir que um dos modelos ajustou melhor em todos os casos, apresentando sempre erros menores que o outro modelo.

**Palavras-chave:** *Epidemiologia, Modelos de transmissão de rumor, Sistemas dinâmicos.*

---

<sup>1</sup>Mestranda do PPGEAB - Unifal-MG. E-mail: *mi\_martins22@hotmail.com*

<sup>2</sup>Mestrando do PPGEAB - Unifal-MG. E-mail: *luiz.marra@hotmail.com*

<sup>3</sup>Professor Doutor do PPGEAB - Unifal-MG. E-mail: *zepaulo@unifal-mg.edu.br*

<sup>4</sup>Agradecimento à FAPEMIG, CAPES, CNPq e UNIFAL pelo apoio financeiro.

# Comparação de testes sobre estruturas de covariância de populações normais

Isabella Marianne Costa Campos<sup>1 4</sup>

Denismar Alves Nogueira<sup>2 4</sup>

Eric Batista Ferreira<sup>3 4</sup>

## 1 Resumo

Em algumas pesquisas existe o interesse em testar a estrutura da matriz de variâncias e covariâncias, como no contexto das diversas técnicas multivariadas ou de modelagem. Diante disto se ressalta a importância da utilização de testes de hipóteses sobre estruturas de covariâncias e assim realizar um estudo detalhado e comparativo de poder e erro tipo I de alguns testes existentes. Foram comparados os testes da razão de verossimilhanças e modificações, considerando-se os cenários com diferentes números de variáveis ( $p = 2, 4, 8, 16, 32, 64$ ) e tamanhos amostrais ( $n = 5, 10, 20, 30, 50, 100$ ). Para avaliar a identidade a proposta de Ledoit e Wolf (2002) foi a mais indicada, para a esfericidade a versão de Jonh (1972) modificada por Ledoit e Wolf (2002) seguida da proposta de Box (1949) foram as de melhor desempenho. No estudo para quaisquer estrutura o teste mais recomendado é uma versão também proposta Ledoit e Wolf (2002).

*Palavras-chave:* Poder; Taxa de erro tipo I; Razão de verossimilhança; Simulação Monte Carlo.

---

<sup>1</sup>DEst - UNIFAL-MG. e-mail: [isabellamccampos@gmail.com](mailto:isabellamccampos@gmail.com)

<sup>2</sup>DEst - UNIFAL-MG. e-mail: [denismar.nogueira@unifal-mg.edu.br](mailto:denismar.nogueira@unifal-mg.edu.br)

<sup>3</sup>DEst - UNIFAL-MG. e-mail: [eric@unifal-mg.edu.br](mailto:eric@unifal-mg.edu.br)

<sup>4</sup>Agradecimento a FAPEMIG pelo apoio financeiro.

## Comparações interculturais de percepção de sabor entre Brasil e Eslováquia utilizando análise de tempo intensidade com intervalos de confiança bootstrap

Fernanda Paes Ribeiro Fernandes Barbosa<sup>1 3</sup>

Isabel de Sousa Amorim<sup>2 3</sup>

Vladimir Vietores<sup>4 3</sup>

Renato Ribeiro de Lima<sup>2 3</sup>

A análise de tempo intensidade (TI) é um método descritivo utilizado na análise sensorial para avaliar a percepção de um atributo ao longo de um período de tempo, com o objetivo de identificar o tempo no qual ocorreu a máxima percepção e avaliar a sua persistência. A análise estatística TI é realizada por meio de curvas, das quais são estimados parâmetros de interesse. Apesar dos parâmetros estimados apresentarem informações relevantes, a maioria dos trabalhos não utilizam intervalos em relação às mesmas, o que possibilita diferenciar estatisticamente diferentes tratamentos ao longo do tempo. O objetivo deste trabalho é utilizar uma nova metodologia com envelopes de confiança para curvas TI, obtidos via simulação *bootstrap*, assim, permitir que inferências sobre as curvas de tempo intensidade sejam realizadas, facilitando e incrementando a análise dos dados. Para ilustrar a metodologia proposta, foram utilizados dados provenientes da análise sensorial de goma de mascar, sabor morango, avaliada por provadores de dois países: Brasil e Eslováquia. Nesse experimento 20 provadores de cada país, entre eles 10 homens e 10 mulheres, avaliaram amostras de goma de mascar em relação ao atributo acidez, ao longo do período de 120 segundos. Os intervalos de confiança propostos foram relevantes e agregaram informações importantes aos resultados da análise de tempo intensidade. As curvas referentes aos intervalos de confiança foram obtidas através do software R. Com as curvas de TI, juntamente com os respectivos intervalos de confiança, foi possível verificar diferenças na sensibilidade de percepção da acidez por indivíduos de diferentes países e gêneros.

**Palavras-chave:** *Bootstrap, Intervalo de confiança, Tempo Intensidade.*

---

<sup>1</sup> Departamento de Ciência dos Alimentos – UFLA. e-mail: fpaesfernandes@gmail.com

<sup>2</sup> Departamento de Estatística – UFLA.

<sup>4</sup> Department of Storing and Processing of Plant Products, Slovak University of Agriculture

<sup>3</sup> Agradecimento a CAPES, CNPq e FAPEMIG pelo apoio financeiro.

## Comportamento da distribuição de renda por tipo de trabalho e nível de instrução dos diferentes grupos sociais no Brasil

Paulo Tadeu Meira e Silva de Oliveira<sup>1 3</sup>

De acordo com o IBGE no censo de 2010, 64,7% da população brasileira tem renda entre 0 e 1 salário mínimo (classe E); 25,8% com renda entre 1 a 3 salários mínimos (classe D); 6,6% com renda entre 3 a 7 salários mínimos (classe C); 2,0% com renda entre 7 a 15 salários mínimos (classe B), e, por fim; 0,9% com renda superior a 15 salários mínimos (classe A). Para os diferentes tipos de trabalho apresentam respectivamente: empregado com carteira assinada (40,0%); militares e funcionários públicos (5,5%); empregados sem carteira assinada (22,0%); conta própria (22,6%); empregadores (1,7%); não remunerados (2,1%), e, por fim; trabalhadores na produção para o próprio consumo (6,1%). Em termos raciais somos constituídos por brancos, 47,0%; negros, 7,1%; amarelos, 1,0%; pardos, 44,3%, e, por fim; indígenas, 0,5%. Resultados que mostra uma distribuição de renda no Brasil profundamente desigual em toda a população. De acordo com a ONU. Além disso, de acordo com especialistas, nota-se, também, desigualdades mais profundas quando considera os diferentes graus de deficiência e de raças cotistas. Neste trabalho, em termos estatísticos fizemos um estudo comparativo da distribuição dos diferentes tipos de trabalho, nível de instrução e de renda em termos proporcionais para os diferentes níveis de deficiência e de raça utilizando técnicas como cruzamentos entre pares de variáveis, teste de homogeneidade, gráfico de análise de perfis e análise de dados composicionais.

**Palavras-chave:** *Pessoas com deficiência, Raça cotista, Teste de homogeneidade, Gráfico de análise de perfis, Análise de dados composicionais*

---

<sup>1</sup> EESC/STT – USP. e-mail poliver@usp.br



## Construção e implementação de modelos de dispersão multivariados

Bruna D. Wundervald <sup>1</sup>

Dennis S. Leão <sup>2</sup>

Wagner Hugo Bonat <sup>3</sup>

Ao construirmos um modelo estatístico, a escolha da distribuição de probabilidades adjacente é crucial. Entretanto, a maioria das distribuições de probabilidade usuais não são adequadas para o tratamento de múltiplas variáveis resposta. A abordagem mais comum é o uso da distribuição normal multivariada, que muitas vezes é inapropriada, uma vez que não reflete o comportamento das variáveis de interesse. Visto isso, este trabalho tem o objetivo de propor distribuições multivariadas para variáveis aleatórias não gaussianas, como binomial, Poisson, normal inversa e gama. As distribuições são construídas de forma que seja possível levar em consideração a estrutura de covariância existente entre as variáveis. O método aplicado emprega formas quadráticas baseadas em resíduos deviance, de forma similar à construção da Normal Multivariada, e foi formulado por Jørgensen and Lauritzen (2000). Dentre os principais resultados, temos a assimilação de como é o comportamento das distribuições não-normalizadas em relação aos parâmetros que as indexam e implementação computacional. Estudos subsequentes envolvem a obtenção e avaliação de aproximações para as constantes normalizadoras, além da construção de modelos de regressão baseados na nova classe de distribuições.

**Palavras-chave:** *distribuições multivariadas, modelos de regressão, resíduos deviance.*

---

<sup>1</sup>LEG/UFPR. e-mail: [brunadaviesw@gmail.com](mailto:brunadaviesw@gmail.com)

<sup>2</sup>LEG/UFPR.

<sup>3</sup>LEG/UFPR.

## Controle estatístico de processo multivariado não paramétrico em uma indústria de laticínios

Luiz Henrique Marra da Silva Ribeiro <sup>1 4</sup>

Tatiane Gomes de Araújo <sup>1 4</sup>

Eric Batista Ferreira <sup>2 4</sup>

Jovelino Zamboni <sup>3 4</sup>

O Brasil é um dos maiores produtores mundiais de leite, e grande parte do leite é destinado à produção queijeira. Para que a indústria de queijo seja mantida, o leite, sua principal matéria-prima deve ser de qualidade, e como as variáveis que determinam as qualidades físico-químicas, microbiológicas e sensoriais do leite são monitoradas de forma contínua na indústria do laticínio, o controle estatístico de processo (CEP) pode ser uma ferramenta para assegurar a qualidade do leite, e por serem várias as variáveis, testes multivariados podem ser uma boa opção. O objetivo desse trabalho foi verificar se o processo de controle de qualidade do leite para a fabricação do queijo de um fabricante do município de São Luís de Montes Belos, Goiás, do ano de 2015 se encontrava sob controle estatístico multivariado, com utilização do teste não paramétrico de postos sinalizados de Wilcoxon, para cinco variáveis do leite: taxa de gordura, extrato seco total (EST), extrato seco desengordurado (ESD), densidade e depressão do ponto de congelamento (DPC). Como testes paramétricos pressupõe conhecimento da distribuição da população, que nem sempre pode ser verificada, uma alternativa é CEP não paramétrico. Com a utilização da técnica, foi verificado que o processo estava sob controle estatístico, pois não foi verificada nenhuma variável fora dos limites de controle, nem sinais de dependência. Porém, estudos futuros são necessários para resultados conclusivos, pois cartas de controle não-paramétricas são menos poderosas e estudos com cartas paramétricas podem ser realizados para fins de confronto e verificação.

**Palavras-chave:** *Inferência não paramétrica, Leite, Queijo, São Luís de Montes Belos.*

---

<sup>1</sup>Mestrando em Estatística Aplicada e Biometria - Unifal. e-mail: [luiz.marra@outlook.com](mailto:luiz.marra@outlook.com); [tatigomesaraujo@yahoo.com.br](mailto:tatigomesaraujo@yahoo.com.br).

<sup>2</sup>Professor Doutor no ICEx - Unifal. e-mail: [eric.ferreira@unifal-mg.edu.br](mailto:eric.ferreira@unifal-mg.edu.br).

<sup>3</sup>Técnico em Laticínios. e-mail: [jozamboni69@yahoo.com.br](mailto:jozamboni69@yahoo.com.br).

<sup>4</sup>Agradecimento à FAPEMIG, à CAPES, ao CNPq e à Unifal pelo apoio financeiro.

## Avaliando genótipos de mostarda com o modelo fator analítico Bayesiano

Carlos Pereira da Silva<sup>134</sup>

Luciano Antonio de Oliveira<sup>234</sup>

Cristian Tiago Erazo Mendes<sup>234</sup>

Joel Jorge Nuvunga<sup>234</sup>

Marcio Balestre<sup>234</sup>

O estudo da interação entre genótipos e ambientes é de suma importância no contexto de programas de melhoramento de plantas e ensaios agrônômicos de uma forma geral. Assim, a utilização de métodos eficientes para analisar dados provenientes de ensaios multiambientes (MET) tem motivado diversas pesquisas, objetivando maior confiança em relação a seleção e recomendação de cultivares superiores. O objetivo deste trabalho foi apresentar uma abordagem Bayesiana para os modelos Fatoriais Analíticos (FA) que são amplamente difundidos nos estudos para avaliar adaptabilidade e estabilidade em dados MET. Um conjunto de dados MET com 12 genótipos de mostarda avaliados em 6 ambientes, os ensaios experimentais foi em delineamentos de blocos casualizados com três repetições. Para analisar os dados MET utilizou-se modelo o fatorial analítico (FA) em uma abordagem Bayesiana. Os resultados evidenciaram flexibilidade na abordagem proposta para identificar genótipos mais produtivos, bem como separar subgrupos de genótipos e ambientes no que se refere ao efeito da interação GEI. Além disso, as representações gráficas para regiões de credibilidade bivariadas das cargas e escores fatoriais, juntamente com regiões de credibilidade bivariadas, ajudaram para um maior detalhamento de GEI, e sendo assim proporcionou mais informação para identificação e recomendação de genótipos.

**Palavras-chave:** *Bayesiana, Modelo Fatorial Analítico, Multiambiente.*

---

1 DES – UFLA. e-mail: ccpsilva81@hotmail.com

2 DES – UFLA.

3 Agradecimento a Capes pelo apoio financeiro.

4 Agradecimento a Fapemig pelo apoio financeiro.

## Estratégias para tratamento de dados incompletos em modelos de regressão logística

José Vitor Frizarin dos Santos<sup>1 3</sup>

Vilma Mayumi Tachibana<sup>2</sup>

Uma das dificuldades no desenvolvimento de modelos preditivos, especificamente o modelo de regressão logística, é a ocorrência de dados *missing*. Os softwares estatísticos mantêm como padrão a análise de dados completos, ou seja, todas as observações com uma ou mais variáveis com valores ausentes são descartadas, reduzindo o tamanho da base de dados que pode resultar em estimativas tendenciosas. Para contornar tal problema, aplicam-se métodos de imputação simples ou múltiplas que estimam os valores ausentes a partir dos dados disponíveis. Neste trabalho comparou-se diferentes métodos de imputação de dados em um estudo do modelo de regressão logística na presença de dados incompletos. Ao todo foram abordados seis métodos de imputação. Um estudo de simulação utilizando dados reais de crédito foi conduzido para comparar o desempenho dos modelos construídos para cada um destes métodos com o modelo de regressão logística padrão ajustado para base de dados completa em dois cenários de dados incompletos. Como esperado, o modelo desenvolvido eliminando as observações com dados faltantes nas variáveis foi o que apresentou o pior desempenho. Já os métodos de imputação *k*-NN e *Random Forest* tiveram a melhor performance, resultando em modelos que apresentam comportamento de predição muito semelhantes ao modelo padrão.

**Palavras-chave:** *Regressão logística, Dados Incompletos, Métodos de Imputação.*

---

<sup>1</sup> FCT - UNESP.

<sup>2</sup> FCT – UNESP. Email: [vilma@fct.unesp.br](mailto:vilma@fct.unesp.br) ou [vilmamayumi@yahoo.com.br](mailto:vilmamayumi@yahoo.com.br)

<sup>3</sup> Agradecimento a Reitoria da Unesp pelo apoio financeiro.

## **Evolução em termos comparativos de pessoas com e sem deficiências no Brasil**

**Paulo Tadeu Meira e Silva de Oliveira<sup>1 3</sup>**

Segundo a OMS em 2011, 1 bilhão de pessoas vivem com alguma deficiência, enquanto que, em termos de Brasil, de acordo com o IBGE, 45,6 milhões são pessoas com deficiência. A falta de estatísticas sobre o tema contribui para a invisibilidade dessas pessoas. Isso representa um obstáculo para planejar e implementar políticas públicas de desenvolvimento que melhorem suas qualidades de vidas. Sabe-se que ter alguma deficiência, de acordo com especialistas, aumenta o custo de vida em cerca de um terço da renda, em média. Completar a escola primária, ou mais ainda, obter um nível de instrução mais avançado é também um desafio maior e tende a crescer quanto pior for o grau de desenvolvimento e assistência do país que reside. Em vista deste cenário, para este trabalho consideramos os dados censitários do IBGE, e, apresentamos um estudo sobre a população classificada para cada ano de idade, as proporções sobre pessoas com deficiência, pessoas que não apresentam deficiência e em conjunto pessoas com e sem deficiência considerando as variáveis estado civil, natureza da união, raça, nível de instrução, renda, tipo de trabalho e condições de moradia. Em termos estatísticos utilizamos gráfico de perfis e dados composicionais.

**Palavras-chave:** *Pessoas com deficiência, Dados censitários, Números proporcionais, Dados composicionais*

---

<sup>1</sup> EESC/STT – USP. e-mail poliver@usp.br

## Ferramenta web para análise de dados culturais usando R e Shiny

Bruno Ogata Franchi<sup>1 3</sup>

Camila Bertini Martins<sup>2 3</sup>

A diversificação do perfil dos estudantes das Instituições Federais de Ensino Superior oriunda da ampliação do acesso é hoje uma realidade. Devido à Lei de Cotas, tem-se atualmente um perfil mais heterogêneo de ingressantes no nível superior. Portanto, mapear o perfil cultural e socioeconômico destes estudantes faz-se necessário para o desenvolvimento de políticas institucionais visando facilitar o diálogo na comunidade acadêmica.

Sabe-se que a Cultura é um dos instrumentos fundamentais na formação de jovens mais tolerantes, generosos, sensíveis e criativos. Nessa perspectiva, exprime-se de cultura como eixo construtor de identidades, como espaço privilegiado de realização da cidadania e de inclusão social e, também, como fator econômico gerador de riquezas. Surge assim, a necessidade da Câmara de Extensão e Cultura da Universidade Federal de São Paulo, campus São José dos Campos, de uma metodologia para trabalhar pontos relacionados à cultura aliado à falta de políticas institucionais ou indicadores que possam nortear o desenvolvimento de propostas de fomento às ações culturais no Instituto.

O objetivo deste trabalho é apresentar uma ferramenta web para facilitar o manuseio da computação estatística na realização de análises de dados culturais da Instituição. A ferramenta foi desenvolvida em linguagem R utilizando o pacote Shiny visando uma interface amigável. Esta interface permite que qualquer pesquisador/servidor realize uma análise estatística de dados culturais de forma rápida, prática e eficiente; subsidiando a elaboração do plano de políticas da Instituição e auxiliando na promoção de cultura, esporte e na permanência estudantil.

**Palavras-chave:** *Dados culturais, Linguagem R, Shiny.*

---

<sup>1</sup> ICT – UNIFESP.

<sup>2</sup> EPM – UNIFESP. e-mail cb.martins@unifesp.br

<sup>3</sup> Agradecimento à PRAE-UNIFESP pelo apoio financeiro.

## Importância da tradução correta de termos técnico-científicos:

### *Traduttore, traditore!*

Alfredo José Barreto Luiz<sup>1</sup>

Termos técnico-científicos são criados em situações específicas, tanto da área do conhecimento quanto da língua em que foram originalmente descritos ou utilizados. Sua tradução para um novo idioma é uma tarefa complexa. No ensino e na aprendizagem é fundamental que um termo guarde coerência com o que significa para tornar mais efetiva a sua compreensão. Isso é importante em todas as áreas do conhecimento, inclusive na Estatística. Exemplos como a tradução de *nugget* para ‘efeito pepita’ em aplicações de geoestatística é um tipo de tradução discutível e limitada. Outro exemplo é traduzir *wavelet* como ‘ondaleta’, uma palavra pouco conhecida e utilizada em português, o que resulta quase o mesmo que não traduzir. Outros exemplos como ‘análise de trilha’ servindo como tradução de *pathway analysis* e *yield gap* sendo vertido para ‘folga de produtividade’, mostram a necessidade de aprimorar a tradução de termos técnico-científicos. Outras vezes a tradução é mesmo errada, como ‘variável categórica’ e não ‘categorizada’ com tradução de *categorical variable* e ‘teorema do limite central’ e não ‘central do limite’ traduzindo *central limit theorem*. Especialistas em terminologia afirmam que definir é expressar um determinado saber ou uma porção de conhecimento especializado. Traduzir uma definição, ou termo, envolve, portanto, uma representação conceitual particular, vinculada a um saber técnico, científico ou tecnológico. O objetivo desse trabalho é ressaltar a importância da tradução correta de termos técnico-científicos em Biometria e Estatística, assim como apresentar exemplos de traduções equivocadas atualmente em uso no Brasil, além de discutir possíveis alternativas mais adequadas.

**Palavras-chave:** *Armadilhas de tradução, terminologia, tradução técnica.*

---

<sup>1</sup> Embrapa Meio Ambiente. E-mail: alfredo.luiz@embrapa.br

## **Imputação de dados na análise de variância em experimentos não balanceados**

**Elisandra Lúcia Moro Stochero<sup>1</sup>**

**Luciane Flores Jacobi<sup>2</sup>**

**Alessandro Dal'Col Lúcio<sup>3</sup>**

**Otávio Flores Jacobi<sup>4</sup>**

**Mateus Bergamo<sup>5</sup>**

É comum deparar-se com eventos não previstos no desenvolvimento de um experimento, havendo muitas vezes a perda de dados. Em um banco de dados completo, os resultados são obtidos sem grandes problemas, porém, quando são desbalanceados, a interpretação das hipóteses pode se tornar difícil. No entanto, existem técnicas de imputação de dados faltantes cujo objetivo é encontrar resultados mais precisos. O objetivo desta pesquisa, que se encontra em andamento, é verificar se há diferença no resultado do teste de Análise de Variância para dados não balanceados quando realizada a inclusão de dados obtidos a partir de uma técnica de imputação de dados. Através de simulações, serão gerados bancos de dados completos, experimento simulado em delineamento inteiramente casualizado com 5 tratamentos e 10 repetições. Posteriormente, serão retirados 10% dos dados de cada banco de dados. Através do Método da Média Preditiva serão obtidas estimativas para os dados que foram retirados e incluídas no banco de dados incompleto. Em cada uma das etapas será realizada a ANOVA para cada banco de dados e posteriormente serão comparados os resultados obtidos determinando a porcentagem de acerto em cada etapa de acordo com os resultados da ANOVA obtidos nos experimentos gerados na primeira etapa.

**Palavras-chave:** *Análise de dados experimentais, ANOVA, Dados faltantes, Métodos de Imputação, Planejamento de experimentos.*

---

<sup>1</sup> BACHARELADO EM ESTATÍSTICA – UFSM. e-mail [elismoro2016@gmail.com](mailto:elismoro2016@gmail.com)

<sup>2</sup> DEPARTAMENTO DE ESTATÍSTICA – UFSM

<sup>3</sup> DEPARTAMENTO DE FITOTECNIA – UFSM

<sup>4</sup> UFRGS

<sup>5</sup> UFSM



## **Maior concorrência no vestibular implica em alunos com maior qualificação?**

**Patrícia Stülz** <sup>1 3</sup>

**Naiara Caroline Ap. dos Santos** <sup>2 3</sup>

**Aline Edlaine de Medeiros** <sup>2 3</sup>

**Jessica Suzana Barragan Alves** <sup>2 3</sup>

**Talita Evelin Nabarrete Tristão de Moraes** <sup>2 3</sup>

Este trabalho visa empregar os conceitos de correlação e regressão linear para avaliar se uma maior concorrência no vestibular está associada a um melhor desempenho dos estudantes aprovados neste exame. Para isso, foram avaliados a nota de corte do último candidato aprovado nos vestibulares de Inverno e Verão da Universidade Estadual de Maringá, e o número de inscritos, no período de 2001-2017, quanto as vagas destinadas ao curso de Estatística. A área de Estatística tem se revelado uma profissão promissora, em concordância com este fato, a análise das informações do vestibular permitiu observar que a busca pela profissão tem aumentado linearmente nos últimos anos. Através do modelo de regressão ajustado nota-se que a nota de corte dos aprovados no vestibular é aproximadamente 7 vezes o número de inscritos, isto é, quanto maior o número de inscritos maior será a nota de corte nos vestibulares para Estatística. O modelo de regressão linear adotado foi avaliado seguindo os critérios de seleção e validação de modelo, apresentando resultado satisfatório. O fato do curso de Estatística estar em ascensão no mercado de trabalho corrobora com os resultados apontados no modelo de regressão, por outro lado, surge a questão de descrever e analisar como tem sido esta dinâmica em áreas que tradicionalmente apresentam maior procura como os cursos de Medicina e Engenharias. Neste sentido, estamos empregando os métodos adotados para avaliar o curso de estatística aos demais curso da Universidade Estadual de Maringá.

**Palavras-chave:** *Regressão Linear, Estatística, Estudantes.*

---

<sup>1</sup>PBE - UEM. e-mail: [patriciastulp2@gmail.com](mailto:patriciastulp2@gmail.com)

<sup>2</sup>PBE - UEM.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

# Maximum entropy distribution on elliptical regions under mean value constraints

J. C. S. de Miranda <sup>1</sup>

Maximum entropy distributions are valuable tools in simulation studies where, in some sense, besides the information we already have about a probability structure, we want to assume the least additional information about it. Using variational methods we determine the maximum entropy probability distribution with support on an elliptical region under mean value constraints. More precisely, we determine the probability density function,  $f_{XY}$ , of a random vector  $(X, Y)$  such that  $\mathcal{Im}(X, Y) \subset \mathcal{E}$ , where  $\mathcal{E} = \{(x, y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\}$ , that maximizes the entropy functional  $f \rightsquigarrow - \int_{\mathcal{E}} f \ln f d\ell$  and satisfies the mean value constraints  $\mathbb{E}X = \mu_X$  and  $\mathbb{E}Y = \mu_Y$ , where  $\mu_X$  and  $\mu_Y$ , such that  $(\mu_X, \mu_Y) \in \mathcal{E}$ , are given.

**Keywords:** *Bessel functions, Elliptical region, Entropy, Variational calculus.*

---

<sup>1</sup>IME - USP. e-mail: *simon@ime.usp.br*

## Meta-analysis tool with R and Shiny

Ricardo Manhães Savii <sup>1 4</sup>

Alexandre Hild Aono <sup>2 4</sup>

Camila Bertini Martins <sup>3</sup>

Although meta-analysis is a general statistics tool for analyzing data from any area of knowledge, its primary application nowadays is with health data. Health professionals, researchers, and health law-makers are flooded by data not always pre-manageable, what makes necessary a systematic review to integrate the data and acquire decisions supporting information. In a systematic review, the data from different studies can be quantitatively combined with statistical methods, called meta-analysis techniques.

Pearson's main argument for using meta-analysis was that most studies are small and do not allow a reliable conclusion by its own. Since then, many authors brought essential considerations and enhanced meta-analysis techniques.

The main goal of this work is to present and explain a web application designed in a way that non-technical users can easily utilize meta-analysis over their data. The web application is fully programmed with the language R version 3.3.1 and its package Shiny as a friendly interface to use most of meta-analysis potential, using R's *meta* package. For the software development, we used IDE R-Studio version 0.99.902 – © 2009-2016 RStudio, Inc, free license.

**Keywords:** *meta-analysis, R programming, shiny.*

---

<sup>1</sup>ICT-UNIFESP - S. J. dos Campos, SP, Brazil. e-mail: [ricardo.manhaes@unifesp.br](mailto:ricardo.manhaes@unifesp.br)

<sup>2</sup>ICT-UNIFESP - S. J. dos Campos, SP, Brazil.

<sup>3</sup>Departamento de Medicina Preventiva, Escola Paulista de Medicina, Universidade Federal de São Paulo - São Paulo, SP, Brazil.

## Modelagem de dados de sobrevivência: uma aplicação a dados de dependentes químicos

Marcelino Alves Rosa de Pascoa <sup>1 3</sup>

Juliany Araujo de Jesus Fernandes de Macedo <sup>1 3</sup>

Juliano Bortolini <sup>1 3</sup>

José Nilton da Cruz <sup>1 3</sup>

Tiago Almeida de Oliveira <sup>2 3</sup>

A distribuição Weibull inversa generalizada (WIG), com três parâmetros tem a capacidade de modelar funções de risco com forma unimodal e decrescente, que são bastante comuns em estudos biológicos e de confiabilidade, e é uma generalização da distribuição Weibull inversa. O objetivo deste trabalho foi através do método de máxima verossimilhança, estimar os parâmetros da distribuição WIG, utilizando dados de dependência química e comparar seu ajuste com o estimador de Kaplan-Meier. Conclui-se que a distribuição Weibull inversa generalizada apresentou um bom ajuste para os dados de dependência química, segundo a comparação com o estimador de Kaplan-Meier.

**Palavras-chave:** *Weibull inversa generalizada, Análise de sobrevivência, Máxima verossimilhança.*

---

<sup>1</sup>DEST - UFMT. e-mail: [marcelino.pascoa@gmail.com](mailto:marcelino.pascoa@gmail.com)

<sup>2</sup>DEST - UEPB.

<sup>3</sup>Agradecimento ICET/UFMT pelo apoio financeiro.

## Modelagem em Extração de Cafeína de sementes de *Paullinia cupana* (guaraná)

Everton da Costa <sup>1 3</sup>

Roney Peterson <sup>2 3</sup>

Vinicius Garcia Tosta <sup>2 3</sup>

Vanderly Janeiro <sup>2 3</sup>

O guaraná (*Paullinia cupana*) é uma planta originária da região amazônica brasileira. São várias as propriedades estimulantes da semente torrada do guaraná tais como aumentar a resistência do organismo e diminuir a fadiga. O Brasil é o maior produtor mundial de guaraná. A produção brasileira tem praticamente sua totalidade consumida no mercado interno, estimando-se que pelo menos 70% da produção seja absorvida pelos fabricantes de refrigerantes.

Esse estudo tem como objetivo verificar quais são as variáveis significativas na extração de cafeína do guaraná. Para cumprir esse objetivo foram utilizados dados coletados na tese de doutorado em ciências farmacêuticas realizada por Marques, L. L. M. . Visando estabelecer a relação entre a quantidade de cafeína extraída e as variáveis regressoras o modelo de regressão linear múltipla se mostrou apropriado. Foram observados o percentual de cafeína extraído e as variáveis tempo (20, 40 e 60 min), temperatura (40, 50 e 60°C), pressão (100, 200 e 300 bar) e modificador (etanol, metanol, etanol/metanol) com percentual de 40%. Foi realizada a seleção de variáveis com base no método *Backward*, *Forward* e *Both* via critério AIC.

Concluiu-se que a cada acréscimo de nível nas variáveis explanatórias Modificador metanol, Tempo de 40 minutos, Tempo de 60 minutos a variável resposta cafeína sofre um decréscimo de 5,37%, 3,91%, 3,52%, respectivamente. Além disso pode-se observar que a cada acréscimo de nível nas variáveis explanatórias Temperatura 60°C e Pressão de 300 bar a variável resposta tem um acréscimo de 5,18% e 1,76%, respectivamente.

**Palavras-chave:** *Modelo de Regressão Linear Múltipla; Stepwise; Região amazônica brasileira.*

---

<sup>1</sup>DES - UEM. e-mail: [everto.cost@gmail.com](mailto:everto.cost@gmail.com)

<sup>2</sup>DES - UEM.

<sup>3</sup>Agradecimento ao Capes pelo apoio financeiro.

## Modelagem temporal de Box & Jenkins da frequência de homicídios na região Sudeste do Brasil durante o período de 2004 à 2017

Rafael Agostinho Ferreira <sup>1 3</sup>  
Patricia Mendes dos Santos <sup>2 3</sup>  
Rodrigo Ferreira de Abreu <sup>2 3</sup>  
João Domingos Scalon <sup>2 3</sup>

Nos últimos anos, a violência no Brasil tem crescido cada vez mais. Diariamente pessoas morrem devido à criminalidade, que tem afetado principalmente os jovens. Dentre as diversas causas de morte, o homicídio doloso figura como uma das principais causas de morte externa. O sudeste tem se mostrado como uma das regiões brasileiras com maior número de casos registrados. A criminalidade e, conseqüentemente, os óbitos ocasionados por este fator são um problema social, o qual precisam ter uma maior atenção por parte dos órgãos voltados à segurança pública. Assim, objetivo deste trabalho foi utilizar a metodologia de séries temporais para modelar a frequência de homicídios na região sudeste do Brasil, durante os anos de 2004 e 2017. Para alcançar os objetivos, foram realizados testes para verificar a presença de tendência e sazonalidade. A partir da série já estacionária, observou-se o comportamento da função de autocorrelação e autocorrelação parcial a fim de propor possíveis modelos aos dados. Baseado em alguns critérios de seleção, tais como o Critério de Informação de Akaike (AIC) e o Erro Quadrático Médio de Previsão (EQMP), constatou-se que o modelo SARIMA(1,1,0)(1,1,0) se adequou bem ao comportamento dos dados. Além disso, foi utilizado o teste de Box-Pierce para testar a hipótese nula de que os resíduos do modelo seguem um ruído branco. Após a escolha do modelo, foi realizada uma previsão para os meses subsequentes. Observou-se que, entre os meses de julho de 2017 e junho de 2018, haverá uma queda no número de ocorrências por tal causa.

**Palavras-chave:** *Criminalidade, Modelo SARIMA, Óbitos.*

---

<sup>1</sup>DES - UFLA. e-mail: [rafaelferreiratc@hotmail.com](mailto:rafaelferreiratc@hotmail.com)

<sup>2</sup>DES - UFLA.

<sup>3</sup>Agradecimento à FAPEMIG e à CAPES pelo apoio financeiro.

## Modelo de regressão polinômio fracionário: uma alternativa aos polinômios convencionais

Edijane Paredes Garcia <sup>1 2</sup>

Luzia Aparecida Trinca <sup>3</sup>

Em modelos de regressão, uma tentativa de representar a não linearidade entre a variável resposta e uma ou mais regressoras são os modelos polinomiais convencionais (PCs). No entanto, PCs de baixa ordem oferecem uma família limitada de curvas e os de alta ordem podem não se ajustar bem a valores extremos. Os modelos de regressão utilizando Polinômios Fracionários (PFs), proposto por Royston & Altman [1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Applied Statistics 43, 429-467], são uma metodologia estatística que combina seleção de modelo com transformação na regressora contínua, oferecendo uma gama de formas funcionais, incluindo a família dos PCs, curvas assimétricas e aquelas com assíntotas, sendo utilizado com sucesso em diversas aplicações. Este trabalho faz do uso dos PFs em modelos com erros normais, mas a metodologia pode ser aplicada aos modelos de Cox, MLG, etc. Para ilustrar a flexibilidade da técnica será apresentado um exemplo, descrito em Salata et al. [2013. Purificação de Etanol de Mandioca com Carvão Ativado. Revista Raízes e Amidos Tropicais 9, 33-41], cujo objetivo foi estudar a condutividade do etanol analisada sob diferentes massas de carvão ativado, com experimento em blocos. Além dos Modelos de Regressão Polinomial e Regressão transformada pelos PFs foi considerado o Modelo Misto PF. Os resultados ilustram algumas das vantagens dos PFs em relação aos PCs que incluem a parcimônia e a flexibilidade da forma da curva. Embora os PFs não sejam necessariamente sugeridos como substitutos para PCs, eles podem ser úteis quando o objetivo é modelar tendências não-lineares com curvas suaves.

**Palavras-chave:** *Curvatura, Forma funcional, Modelo misto, Modelo polinomial fracionário.*

---

<sup>1</sup> ICE - UFAM. e-mail: [edijane@ufam.edu.br](mailto:edijane@ufam.edu.br)

<sup>2</sup> PGG Biometria - UNESP. e-mail: [edijane@ibb.unesp.br](mailto:edijane@ibb.unesp.br)

<sup>3</sup> IBB - UNESP. e-mail: [ltrinca@ibb.unesp.br](mailto:ltrinca@ibb.unesp.br)

# Modelos de séries temporais para análise de óbitos de motoristas e passageiros de automóveis em Minas Gerais

Patricia Mendes dos Santos <sup>1 3</sup>

Rafael Agostinho Ferreira <sup>2 3</sup>

Vânia de Fátima Lemes de Miranda <sup>2 3</sup>

Thelma Sáfadi <sup>2 3</sup>

Marcelo Ângelo Cirillo <sup>2 3</sup>

A análise de séries temporais é muito utilizada para analisar observações ordenadas no tempo. Através desta análise é possível identificar padrões não aleatórios na série temporal de uma variável de interesse, e a observação deste comportamento passado pode permitir fazer previsões sobre o futuro, orientando na tomada de decisões. Sua principal diferença entre os demais métodos de análise está no fato de considerar e modelar a correlação presente nos erros. Desse modo, este trabalho tem como objetivo utilizar técnicas de séries temporais para analisar o número de óbitos de motoristas e passageiros de automóveis, ocorridos em acidentes de trânsito no estado de Minas Gerais, entre os meses de janeiro de 1996 à janeiro de 2015. Para esta análise, foram verificadas a variabilidade da série, assim como a presença das componentes tendência e sazonalidade. A metodologia Box-Jenkins foi utilizada na modelagem dos dados, sendo que a série em estudo foi bem ajustada utilizando modelos da classe ARIMA. A comparação entre os modelos considerados adequados à série foi realizada através do AIC e EQMP. Entre os modelos sugeridos, o modelo ARIMA(2,1,0) com 8 intervenções foi considerado o modelo mais adequado para descrever comportamento da série. O modelo ajustado foi utilizado para fazer previsões sobre as observações futuras dessa série. Além disso, ajustou-se um modelo ARCH(1) para a volatilidade presente na série.

**Palavras-chave:** *Box-Jenkins, Modelo ARIMA, Óbitos, Previsão.*

---

<sup>1</sup>DES - UFLA. e-mail: [patymendesdosantos@hotmail.com](mailto:patymendesdosantos@hotmail.com)

<sup>2</sup>DES - UFLA.

<sup>3</sup>Agradecimento à FAPEMIG e à CAPES pelo apoio financeiro.



# Modelos duplos COM-Poisson: modelando média e dispersão na análise de contagens

Eduardo Elias Ribeiro Junior <sup>† 1 2</sup>

Clarice Garcia Borges Demétrio <sup>1</sup>

Para análise de dados em forma de contagens, comumente, modela-se a média da variável resposta em termos de covariáveis. A relação média-variância é determinada ao especificar a distribuição de probabilidades das contagens. Dessa forma, a variância das contagens se relaciona com as covariáveis apenas por meio de suas médias. Nesse artigo, propõem-se os modelos duplos COM-Poisson, em que adota-se a distribuição COM-Poisson para as contagens e modela-se ambos os parâmetros da distribuição, de média e de dispersão, em termos de covariáveis. Essa abordagem garante flexibilidade suficiente para identificar covariáveis influentes tanto na média quanto na dispersão das contagens em casos de sub, equi e superdispersão. Os modelos propostos são ajustados pelo método da máxima verossimilhança e a inferência sob os parâmetros é baseada na distribuição assintótica dos estimadores de máxima verossimilhança. A metodologia é aplicada para análise de um estudo em biometria. Os resultados dos modelos duplos COM-Poisson apresentaram melhorias em termos de ajuste e interpretação quando comparados aos resultados dos modelos COM-Poisson convencionais. As implementações computacionais para ajuste dos modelos duplos COM-Poisson são disponibilizadas em material suplementar.

**Palavras-chave:** *Distribuição COM-Poisson, Modelagem da dispersão, Modelos duplos COM-Poisson.*

---

<sup>†</sup>Contato: [jreduardo@usp.br](mailto:jreduardo@usp.br)

<sup>1</sup>Departamento de Ciências Exatas (LCE) - ESALQ-USP

<sup>2</sup>Laboratório de Estatística e Geoinformação (LEG) - UFPR

## **Modelos para dados categorizados ordinais com efeito aleatório: uma aplicação à análise sensorial**

**Maíra Blumer Faretto<sup>1 3</sup>**

**Idemauro Antonio Rodrigues de Lara<sup>2</sup>**

Os Modelos de Logitos Cumulativos, em que a função de ligação é constituída de probabilidades acumuladas, são muito utilizados dados categorizados ordinais, sendo uma de suas simplificações, os Modelos de Chances Proporcionais, em que para todas as covariáveis no modelo há um crescimento linear nas razões de chances. Outros modelos como o Modelo de Chances Proporcionais Parciais, o Modelo de Categorias Adjacentes e o Modelo Logito de Razão Contínua também podem ser utilizados. Em diversos estudos deste tipo, é necessário a utilização de modelos mistos, seja pelo tipo de um fator ou a dependência entre observações da variável resposta. Objetivou-se, neste trabalho, o estudo de modelos para variável resposta ordinal com a inclusão de um ou mais efeitos aleatórios. Esses modelos são ilustrados com a utilização de dados reais de análise sensorial, cuja variável resposta é constituída de uma escala ordinal e deseja-se saber dentre duas variedades de tomates desidratados (Italiano e Sweet Grape), qual teve melhor aceitação pelos consumidores. Nesse experimento os provadores avaliaram uma única vez cada uma das variedades, sendo as repetições constituídas pelas avaliações dadas por diferentes provadores. Nesse caso, é necessária a inclusão de um efeito aleatório por provador, para que o modelo consiga capturar as diferenças entre esses provadores não treinados. O Modelo de Chances Proporcionais ajustou-se de maneira satisfatória aos dados, podendo-se fazer uso das estimativas de probabilidades e razões de chances para a interpretação dos resultados e concluindo-se que o sabor da variedade Sweet Grape foi o que mais agradou os provadores.

**Palavras-chave:** *Dados categorizados; Modelos de logitos cumulativos; Modelos lineares generalizados mistos.*

---

<sup>1</sup> ESALQ – USP. e-mail: mairafaretto@gmail.com

<sup>2</sup> ESALQ – USP.

<sup>3</sup> Agradecimento ao CNPq pelo apoio financeiro.

## Níveis de retorno para valores máximos de temperatura de Piracicaba-SP

Valdeline de Paula Mequelino Ferreira <sup>1 2 4</sup>

Luiz Alberto Beijo <sup>3 4</sup>

Fabricao Goecking Avelar <sup>3 4</sup>

A frequência e a intensidade dos eventos extremos climáticos estão aumentando. Por essas razões é cada vez maior a necessidade de se prever estas ocorrências. A Distribuição Generalizada de Valores Extremos (GEV) e a Distribuição Gumbel são comumente usadas no estudo de eventos extremos. O presente trabalho teve como objetivo ajustar estas distribuições para os dados de temperatura máxima mensal compreendidos entre os anos de 1917 a 2018, obtidos junto à Estação Convencional do Posto Meteorológico da Escola de Agricultura Luiz de Queiroz (ESALQ/USP), em Piracicaba-SP. A independência dos dados mensais de temperatura máxima foi verificada por meio do teste de Ljung-Box e a estacionalidade pelo teste de Mann-Kendal, ambos ao nível de 5% de significância. A qualidade do ajuste foi verificada por meio dos gráficos de PP-Plot e QQ-Plot. Para cada mês foram preditos os níveis de retorno para os tempos de retorno de 5, 10, 20, 50 e 100 anos. Os dados de temperatura máxima mensal para os meses de janeiro, maio, junho, novembro e dezembro não apresentaram tendência. A Distribuição Gumbel foi ajustada para o mês de dezembro e, para os demais meses, foi ajustada a Distribuição GEV. Os dados dos meses de fevereiro, março, abril, julho, agosto, setembro e outubro apresentaram tendência. Nesses casos, foram ajustadas a Distribuição Gumbel com tendência para o mês de fevereiro e a distribuição GEV com tendência para os outros meses. Concluiu-se que a temperatura máxima mensal de Piracicaba-SP está aumentando para a maioria dos meses.

**Palavras-chave:** *Eventos extremos, Distribuição GEV, Distribuição Gumbel.*

---

<sup>1</sup>Graduanda em Matemática - UNIFAL-MG. e-mail: [valdelinedepaula@hotmail.com](mailto:valdelinedepaula@hotmail.com)

<sup>2</sup>Bolsista CNPq

<sup>3</sup>Instituto de Ciências Exatas - UNIFAL-MG.

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

# Modelo Não-Gaussiano de Volatilidade Estocástica com *Jumps* via Gibbs Sampler

Arthur Tarso Rego<sup>1 3</sup>

Thiago Rezende dos Santos<sup>2 3</sup>

Compreender o comportamento dos preços dos ativos é essencial para decisões de alocação de capital entre a opção de investimento disponível. Essa decisão depende do que se pensa sobre riscos e retornos associados a essas opções de investimento. Na literatura, existem modelos focados na estimativa de risco de ativos financeiros, no entanto, a maioria deles depende de métodos MCMC com base em algoritmos Metropolis, o que os torna computacionalmente caros. Este trabalho apresenta um modelo alternativo capaz de estimar a volatilidade sem recurso a métodos computacionais intensivos. Uma análise com dados reais é realizada e os resultados indicam que o modelo proposto foi capaz de estimar a volatilidade, de forma computacional eficiente e automatizada.

**Palavras-chave:** *Gibbs Sampler, Modelos de Volatilidade Estocástica, Modelos Dinâmicos, Séries Temporais.*

---

<sup>1</sup>DEST - UFMG. e-mail: [arthur.tarso@outlook.com](mailto:arthur.tarso@outlook.com)

<sup>2</sup>DEST - UFMG. e-mail: [thiagords@est.ufmg.br](mailto:thiagords@est.ufmg.br)

<sup>3</sup>Agradecimento à FAPEMIG e UFMG pelo apoio financeiro.

## O argumento de encolhimento

Tiago Moreira Vargas <sup>1 3</sup>

O argumento de encolhimento é uma técnica bayesiana para a obtenção de esperança de variáveis aleatórias. Seja  $\mathbf{x} = (x_1, \dots, x_n)^\top$  um vetor aleatório que possui função densidade de probabilidade  $f(\cdot; \boldsymbol{\theta})$  que depende de um vetor de parâmetros  $p$ -dimensional  $\boldsymbol{\theta} \in \Theta$ , onde  $\Theta \subseteq \mathbb{R}^p$  é um subconjunto aberto do espaço Euclidiano. Tal técnica foi criada por Ghosh e Mukerjee (1991), e consiste em uma rota bayesiana bem definida em três passos, a fim de se obter a esperança de uma função de interesse  $Q(\mathbf{x}, \boldsymbol{\theta})$ , isto é,  $E_{\boldsymbol{\theta}}(Q(\mathbf{x}, \boldsymbol{\theta}))$ .

Este trabalho descreve minuciosamente a técnica do argumento de encolhimento, e desenvolve exemplos práticos de sua utilização em situações de interesse, a saber: correção de viés do estimador de máxima verossimilhança e desenvolvimento de regiões de credibilidade acuradas.

**Palavras-chave:** *rota bayesiana, argumento de encolhimento.*

---

<sup>1</sup>Instituto de Matemática e Estatística, Universidade Federal de Goiás. e-mail: *tiagomvargas@gmail.com*

## On the prediction error

Leandro da Silva Pereira<sup>1 4</sup>

Lucas Monteiro Chaves<sup>2 4</sup>

Devanil Jaques de Souza<sup>3 4</sup>

A statistical model should, like almost every scientific procedure, have one eye on the past and two eyes on the future. Once a model has been fitted to a data set, how well it will describe future observations? In textbooks, generally, little attention is given on measuring the capacity of these models to predict future data. This work intends to develop a trivial, yet exhaustive, approach explaining all the mathematical passages in order to facilitate the reader's understanding on the prediction error of a model, having been motivated by the first two pages in Efron(2004). We use an intuitive tool in order to resume the question on triangles and construct a simple way to visualize the geometry behind the prediction error described by the classical paper of Efron(2004).

As a contrast to the geometric approach, the algebraic development related with the theory is presented once both of them complement each other. Two elementary examples are presented: the first one is the simplest case where the model is just the mean of the data. The second one describes the prediction error when the least-squares estimator is taken as the model.

**Palavras-chave:** *model fitting, double expectation, covariance, geometry.*

---

<sup>1</sup>DAMAT - UTFPR. e-mail: [leandropereira@utfpr.edu.br](mailto:leandropereira@utfpr.edu.br)

<sup>2</sup>DEX - UFLA. e-mail: [lucas@dex.ufla.br](mailto:lucas@dex.ufla.br)

<sup>3</sup>DES - UFLA. e-mail: [devaniljaques@des.ufla.br](mailto:devaniljaques@des.ufla.br)

<sup>4</sup>Thanks to CNPq for the financial support.

## Propriedades multifractais da umidade relativa no Nordeste brasileiro

Hérica Santos da Silva<sup>1,3</sup>

José Rodrigo Santos Silva<sup>2</sup>

Tatijana Stosic<sup>1</sup>

A umidade relativa é uma importante variável climática que tem efeito direto sobre a precipitação e, portanto, o conhecimento da dinâmica da umidade relativa é de extrema importância. O nordeste brasileiro, devido sua grande extensão e localização é influenciado pelos vários sistemas atmosféricos e possui clima semiárido caracterizado pelos baixos níveis pluviiais com alta variação temporal e espacial, temperaturas elevadas e altas taxas de evaporação. A fim de contribuir para uma melhor compreensão do processo estocástico subjacente que gera a variabilidade temporal da umidade relativa do ar no Nordeste do Brasil, aplicamos o método de análise de flutuação destendenciada para multifractais (MF-DFA) para analisar as propriedades multifractais de séries históricas desta variável, fornecidas pelo Instituto Nacional de Meteorologia (INMET) em 94 estações meteorológicas distribuídas na região. No Nordeste brasileiro, a umidade relativa média foi de 70,6% com uma variabilidade 8,5%. Verificou-se que as series de anomalias da umidade relativa possuem propriedades multifractais. Além disso, a partir dos parâmetros multifractais extraídos do espectro de singularidade (largura do espectro ( $\Delta\alpha$ ), assimetria ( $\Delta_{Ass}$ ) e o expoente de correlação de longo alcance ( $\alpha_0$ )), observou-se que em todo o Nordeste brasileiro observou-se que os valores  $\alpha_0$  são maiores que 0,5 indicando persistência climática para a umidade relativa do ar. A largura do espectro variou de 0,15 a 0,48 com largura média  $\Delta\alpha = 0,27 \pm 0,05$ . Os

---

<sup>1</sup> DEINFO – UFRPE. e-mail hericasilva58@hotmail.com

<sup>2</sup> DECAT – UFS.

<sup>3</sup> Agradecimento ao Capes pelo apoio financeiro.

valores do parâmetro de assimetria ( $\Delta_{ASS}$ ) foram todos maiores que 1, revelando que a umidade relativa em todo o Nordeste é mais influenciado por pequenas flutuações.

**Palavras-chave:** *MF-DFA, nordeste, umidade relativa.*



*Revisitando o Gráfico Biplot proposto por Gabriel (1971).**1. Partição da Soma de Quadrados.*Luiz Roberto Martins Pinto<sup>1</sup>,Carlos Tadeu dos Santos Dias<sup>2</sup>

Resumo: Gabriel (1971) afirma que “The biplot provides a useful tool of data analysis and allows the visual appraisal of the structure of large data matrices. It is especially revealing in principal component analysis, where the biplot can show inter-unit distances and indicate clustering of units as well as display variances and correlations of the variables”. Ainda que a análise realizada por meio do Gráfico Biplot (GB) seja, conceitualmente, de valor indiscutível, o método analítico sugerido para o cálculo das suas coordenadas, e para a fatoração (ou partição) de  $\lambda$ , em função dos vetores singulares  $p$  e  $q$  não atende aos objetivos propostos. Aqui nós propomos a partição de  $\lambda^2$  para representar a variabilidade por meio do GB, e que as coordenadas do GB sejam calculadas em função de  $p^2$  e  $q^2$ .

Palavras-chave: Biplot, coordenadas, fatoração.

---

<sup>1</sup> UESC e-mail: [luizroberto.uesc@gmail.com](mailto:luizroberto.uesc@gmail.com)

<sup>2</sup> USP-ESALQ: e-mail: [ctsdias@usp.br](mailto:ctsdias@usp.br)

*Revisitando o Gráfico Biplot proposto por Gabriel (1971).**2. Partição do Quadrado Médio.*Luiz Roberto Martins Pinto<sup>1</sup>,Carlos Tadeu dos Santos Dias<sup>2</sup>

Resumo: O gráfico Biplot sugerido por Gabriel (1971) é construído à partir de coordenadas resultantes da fatoração (ou partição) de  $\lambda$ , em função dos autovetores  $p$  e  $q$ . Aqui nós propomos a fatoração de  $\sigma_k^2 = \lambda_k^2 / gl_k$  (em que  $\lambda_k^2$ ,  $gl_k$  e  $\sigma_k^2$  são, respectivamente, a soma de quadrados, o grau de liberdade e a variância associados a cada um  $k$  dos componentes principais (CP's)). A principal vantagem nesta fatoração é a avaliação mais adequada dos CP's a serem retidos para a explicação da variabilidade padrão da matriz em questão, posto que a soma de quadrados associada a cada um deles é ponderada por seus respectivos  $gl$ 's, o que pode, por exemplo, ensejar a criação de novos testes probabilísticos para comparações entre os níveis dos fatores envolvidos.

Palavras-chave: Biplot, coordenadas, fatoração do quadrado médio.

---

<sup>1</sup> UESC e-mail: [luizroberto.uesc@gmail.com](mailto:luizroberto.uesc@gmail.com)

<sup>2</sup> USP-ESALQ: e-mail: [ctsdias@usp.br](mailto:ctsdias@usp.br)

## Robustez de uma adaptação do teste $T^2$ de Hotelling para o vetor de médias sob distribuições não normais

Henrique José de Paula Alves<sup>1 3</sup>  
Daniel Furtado Ferreira<sup>2 3</sup>

Na análise estatística em geral, o pressuposto de normalidade é essencial para a realização de diversas inferências em muitos casos. A presença de “outliers” nas observações pode afetar a tomada de decisão quando realizamos algum teste de hipótese, em especial o teste de médias. A taxa de erro Tipo I e o poder do teste podem ficar comprometidos. Na análise multivariada, em especial no caso bivariado, uma estatística de teste para o vetor de médias muito utilizada é a estatística  $T^2$  de Hotelling tradicional, que leva em conta a distância de Mahalanobis entre as observações e a matriz de covariâncias, que é desconhecida e, portanto, deve ser estimada. Porém na presença de “outliers” e distribuições não normais esta pode não ser a melhor escolha uma vez que o teste resultante pode sofrer aumento na taxa de erro Tipo I e perda de poder. Como vários métodos para estimação robusta existem, pretende-se utilizar o estimador introduzido por Falk (1997) que é baseado em medianas correlacionadas (COMEDIAN), sendo robusto à presença de “outliers”. Neste trabalho, pretende-se avaliar o desempenho de um novo teste usando-se uma adaptação da estatística  $T^2$  de Hotelling que incorpore esse estimador avaliando-se a taxa de erro Tipo I e o poder do teste.

**Palavras-chave:** *COMEDIAN,  $T^2$  Hotelling, outliers.*

---

<sup>1</sup> DES – UFLA. e-mail [hjpadex@posgrad.ufla.br](mailto:hjpadex@posgrad.ufla.br)

<sup>2</sup> DES – UFLA, Bolsista do CNPq.

<sup>3</sup> Agradecimento ao CNPq, à FAPEMIG e à CAPES pelo apoio financeiro.

## Seleção de covariáveis com regressão LASSO e Elastic Net

Carlos José dos Reis <sup>1 3</sup>

Laerte Dias de Carvalho <sup>2 3</sup>

Lucas Monteiro Chaves <sup>4</sup>

Devanil Jaques de Souza <sup>5</sup>

Uma regressão contendo diversas covariáveis regressoras torna o modelo muito complexo e ocasiona a perda ou a dificuldade na interpretação dos resultados. Uma maneira de eliminar esse problema é realizar a retenção de um subconjunto de coeficientes regressores que não sejam significativos para a análise do fenômeno em estudo. Dentre os métodos de encolhimento, denominados de métodos *Shrinkage*, utilizados para essa finalidade, podem-se citar os métodos *Ridge*, *LASSO* (do inglês “*Least Absolute Shrinkage and Selection Operator*”) e *Elastic Net*. Diante disso, objetivou-se no presente trabalho avaliar esses três métodos utilizando um conjunto de  $p = 50$  covariáveis simuladas, cada uma contendo  $n = 100$  observações. O modelo utilizado na simulação foi  $y = X\beta + \varepsilon$ , em que  $\beta = \left( 10, 10, 5, 5, 5, \underbrace{1, \dots, 1}_{10}, \underbrace{0, \dots, 0}_{35} \right)$  e  $\varepsilon$  é Normal padrão. A correlação entre variáveis regressoras  $X_i$  e  $X_j$  foi dada por  $0,5^{|i-j|}$ . Para avaliar a acurácia dos métodos utilizou-se o critério do menor erro quadrático médio (*EQM*). Como era esperado, pode-se verificar pelos resultados que o pior desempenho em termos de *EQM* foi observado para o método *Ridge* ( $EQM = 12,43$ ). Esse resultado deve-se ao fato da regressão de *Ridge* não selecionar covariáveis, principalmente numa situação de muitas variáveis preditoras. A melhor performance foi observada para o método *LASSO* ( $EQM = 1,77$ ), seguido pela metodologia *Elastic Net* ( $EQM = 1,79$ ). Para o presente cenário de simulação, as metodologias *LASSO* e *Elastic Net* são indicadas, uma vez que ambas apresentaram valores de *EQM* muito próximos.

**Palavras-chave:** *Acurácia, Encolhimento, Erro Quadrático Médio.*

<sup>1</sup>DES - UFLA. e-mail: [carlosjreis17@yahoo.com.br](mailto:carlosjreis17@yahoo.com.br)

<sup>2</sup>DES - UFLA. e-mail: [laerte@ufv.br](mailto:laerte@ufv.br)

<sup>3</sup>Agradecimentos à CAPES, ao CNPq e à FAPEMIG pelo apoio financeiro.

<sup>4</sup>DEX - UFLA. e-mail: [lucas@dex.ufla.br](mailto:lucas@dex.ufla.br)

<sup>5</sup>DES - UFLA. e-mail: [devaniljaques@des.ufla.br](mailto:devaniljaques@des.ufla.br)

## Seleção de covariáveis com regressão LASSO e Elastic Net

Carlos José dos Reis <sup>1 3</sup>

Laerte Dias de Carvalho <sup>2 3</sup>

Lucas Monteiro Chaves <sup>4</sup>

Devanil Jaques de Souza <sup>5</sup>

Uma regressão contendo diversas covariáveis regressoras torna o modelo muito complexo e ocasiona a perda ou a dificuldade na interpretação dos resultados. Uma maneira de eliminar esse problema é realizar a retenção de um subconjunto de coeficientes regressores que não sejam significativos para a análise do fenômeno em estudo. Dentre os métodos de encolhimento, denominados de métodos *Shrinkage*, utilizados para essa finalidade, podem-se citar os métodos *Ridge*, *LASSO* (do inglês “*Least Absolute Shrinkage and Selection Operator*”) e *Elastic Net*. Diante disso, objetivou-se no presente trabalho avaliar esses três métodos utilizando um conjunto de  $p = 50$  covariáveis simuladas, cada uma contendo  $n = 100$  observações. O modelo utilizado na simulação foi  $y = X\beta + \varepsilon$ , em que  $\beta = \left( 10, 10, 5, 5, 5, \underbrace{1, \dots, 1}_{10}, \underbrace{0, \dots, 0}_{35} \right)$  e  $\varepsilon$  é Normal padrão. A correlação entre variáveis regressoras  $X_i$  e  $X_j$  foi dada por  $0,5^{|i-j|}$ . Para avaliar a acurácia dos métodos utilizou-se o critério do menor erro quadrático médio (*EQM*). Como era esperado, pode-se verificar pelos resultados que o pior desempenho em termos de *EQM* foi observado para o método *Ridge* ( $EQM = 12,43$ ). Esse resultado deve-se ao fato da regressão de *Ridge* não selecionar covariáveis, principalmente numa situação de muitas variáveis preditoras. A melhor performance foi observada para o método *LASSO* ( $EQM = 1,77$ ), seguido pela metodologia *Elastic Net* ( $EQM = 1,79$ ). Para o presente cenário de simulação, as metodologias *LASSO* e *Elastic Net* são indicadas, uma vez que ambas apresentaram valores de *EQM* muito próximos.

**Palavras-chave:** *Acurácia, Encolhimento, Erro Quadrático Médio.*

<sup>1</sup>DES - UFLA. e-mail: carlosjreis17@yahoo.com.br

<sup>2</sup>DES - UFLA. e-mail: laerte@ufv.br

<sup>3</sup>Agradecimentos à CAPES e a FAPEMIG pelo apoio financeiro.

<sup>4</sup>DEX - UFLA. e-mail: lucas@dex.ufla.br

<sup>5</sup>DES - UFLA. e-mail: devaniljaques@des.ufla.br

*Seleção de objetos e de variáveis em grandes matrizes de dados.*

Luiz Roberto Martins Pinto<sup>1</sup>,  
Carlos Tadeu dos Santos Dias<sup>2</sup>

Resumo: Um dos grandes problemas na análise estatística de grandes matrizes de dados (e.g. genômica) é a seleção das variáveis e dos objetos de estudo que representem a variabilidade padrão no fenômeno em pesquisa. Aqui nós propomos uma metodologia para a seleção de variáveis e de objetos com fundamentos na análise de componentes principais e análise biplot.

Palavras-chave: Seleção de objetos, Seleção de variáveis, ACP, Biplot.

---

<sup>1</sup> UESC e-mail: [luizroberto.uesc@gmail.com](mailto:luizroberto.uesc@gmail.com)

<sup>2</sup> USP-ESALQ: e-mail: [ctsdias@usp.br](mailto:ctsdias@usp.br)

## Teste de normalidade multivariada baseado em amostras da distribuição F de Snedecor

Renata Aparecida Cintra <sup>1 4</sup>

Daniel Furtado Ferreira <sup>2 4</sup>

Marcel Irving Pereira Mélo <sup>3 4</sup>

A verificação de normalidade multivariada é muito importante, pois diversos métodos estatísticos são baseados em hipóteses de que os dados provêm de uma distribuição normal multivariada. Consequentemente, o uso de um teste de hipótese é indispensável para presumir a normalidade. Ademais, não há, na literatura, um único teste que seja o mais poderoso em todas as situações. Este trabalho teve como principal objetivo propor um teste para normalidade multivariada com base em uma amostra univariada da distribuição F de Snedecor, obtida, mediante transformações em uma amostra aleatória normal  $p$ -variada. As estimativas das taxas de erro tipo I e o poder do teste foram obtidas a partir de simulações Monte Carlo. Foram geradas amostras de diferentes tamanhos  $n$  e diferentes dimensões  $p$  da distribuição normal  $p$ -variada e de outras distribuições não normais. O teste de normalidade multivariada Shapiro-Wilk foi avaliado para fim de comparação. Conclui-se que o teste proposto apresentou bom controle das taxas de erro tipo I, sendo um teste exato. Todavia, em relação ao poder, esse teste não obteve bom desempenho. Portanto, a sua utilização não é recomendada.

**Palavras-chave:** *Distribuição F, Kolmogorov-Smirnov, Método Monte Carlo.*

---

<sup>1</sup>Departamento de Ciências Exatas - UFLA. e-mail: [renata\\_cintra\\_spu@hotmail.com](mailto:renata_cintra_spu@hotmail.com)

<sup>2</sup>Departamento de Estatística - UFLA. email: [danielff@des.ufla.br](mailto:danielff@des.ufla.br)

<sup>3</sup>Departamento de Estatística - UFLA. e-mail: [mipmelo90@gmail.com](mailto:mipmelo90@gmail.com)

<sup>4</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

# Teste de Normalidade Univariada Baseado na Entropia

Neto Pascoal <sup>1 3</sup>

Daniel Furtado Ferreira <sup>2 3</sup>

O comportamento de muitos fenômenos da vida real em muitas áreas do saber é descrito pelas distribuição normal de probabilidade. Ao se retirar uma amostra aleatória de uma população, no caso univariado, é comum pressupor que os dados sejam provenientes de uma população normal. Essa pressuposição de normalidade deve ser verificada por meio da aplicação de testes estatísticos sobre a normalidade. Embora a entropia seja entendida como uma medida da quantidade de aleatoriedade de um sistema de informação, sendo usada para medir a incerteza de uma variável aleatória, ela tem sido utilizada para testar a normalidade. Este trabalho tem como objetivo propor testes de normalidade univariada baseados na entropia. Estes testes consistem em aplicar procedimentos computacionalmente intensivos baseados na técnica *bootstrap* paramétrico. Além disso, pretende-se comparar o poder do teste de normalidade Shapiro-Wilk com os testes de normalidade univariada propostos. Os testes propostos obtiveram um controle adequado das taxas de erro tipo I, se mostrando exatos com tamanhos iguais ao nível de significância nominal  $\alpha$ .

**Palavras-chave:** *Bootstrap paramétrico, poder; erro tipo I, teste de Shapiro-Wilk.*

---

<sup>1</sup>DES - Depto de Estatística - UFLA, e-mail: [polenepascoal@gmail.com](mailto:polenepascoal@gmail.com)

<sup>2</sup>DES - Depto de Estatística- UFLA, Caixa Postal 37 e-mail: [danielff@des.ufla.br](mailto:danielff@des.ufla.br)

<sup>3</sup>Agradecimento ao CNPq e ao IBE-Moçambique pelo apoio financeiro.



## Testes de Hipóteses *fuzzy* aplicados em Ciências Atuariais

William Oliveira Santos <sup>1 5</sup>

Leandro Ferreira <sup>2 5</sup>

Danilo Machado Pires <sup>3 5</sup>

Silvio Antônio Bueno Salgado <sup>4 5</sup>

Ao se realizar um teste de hipótese estatístico, pode-se encontrar situações em que as hipóteses são analisadas em ambiente de incerteza *fuzzy*, onde termos linguísticos como “aproximadamente” e “em torno de” podem ser considerados nas definições das hipóteses. O objetivo deste trabalho é apresentar aplicações de testes de hipóteses *fuzzy* em Ciências Atuariais, utilizando conjuntos *fuzzy* representados por funções de pertinência.

**Palavras-chave:** *Matemática Atuarial, Estatística fuzzy, Hipótese fuzzy.*

---

<sup>1</sup>UNIFAL-MG. e-mail: *william.osantos96@gmail.com*

<sup>2</sup>UNIFAL-MG. e-mail: *leandro.ferreira@unifal-mg.edu.br*

<sup>3</sup>UNIFAL-MG. e-mail: *danilo.pires@unifal-mg.edu.br*

<sup>4</sup>UNIFAL-MG. e-mail: *silvio.salgado@unifal-mg.edu.br*

<sup>5</sup>Agradecimento a FAPEMIG pelo apoio financeiro.

## Uma análise das variáveis que impactam na quantidade de massa gorda corporal em indivíduos que praticam natação

Naiara Caroline Ap. dos Santos <sup>1 3</sup>

Patrícia Stülp <sup>2 3</sup>

Aline Edlaine de Medeiros <sup>2 3</sup>

Terezinha Aparecida Guedes <sup>2 3</sup>

Roney Peterson Pereira <sup>2 3</sup>

Uma escola de natação de Maringá, recebeu 237 novos usuários no ano de 2017. Tais usuários buscaram esta atividade com o intuito de melhorar sua condição física. A escola avaliou os indivíduos quanto a sua composição corporal, incluindo análise da obesidade, a proporção de músculo-gordura, entre outros. Neste estudo, coletaram-se as observações corporais destes indivíduos com a finalidade de analisar se as diversas características corporais podem explicar a massa gorda corporal do indivíduo. Após a realização de uma análise descritiva dos dados, foram avaliados alguns modelos que poderiam descrever a relação entre tais variáveis e assim sendo, notou-se que o modelo de regressão linear múltipla tem suas condições de aplicação satisfeitas para as variáveis observadas, bem como atende os critérios de seleção de modelo. De acordo com o modelo, observou-se que as variáveis quantidade de minerais por indivíduo, massa do músculo esquelético e índice de massa corporal explicam a massa gorda corporal do indivíduo. Em particular, quanto maior é o índice de massa corporal e quantidade de minerais do sujeito, maior será a quantidade de massa gorda corporal apresentada por ele. A medida em que o sujeito eleva sua massa magra, por exemplo, praticando natação, a sua massa gorda corporal tende a diminuir. O conhecimento dessas informações pode ser útil para a escola e para os sujeitos no desenvolvimento de atividades específicas para redução da massa corporal gorda.

**Palavras-chave:** *Regressão Linear Múltipla, Estatística.*

---

<sup>1</sup>PBE - UEM. e-mail: [naicaroline2@gmail.com](mailto:naicaroline2@gmail.com)

<sup>2</sup>PBE - UEM.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

## Uma aplicação do modelo de resposta aleatorizada

Olga Lidia Solano Dávila<sup>1 5</sup>

Jose Ernaú Huaman<sup>2 5</sup>

Felix Bartolo Gotarate<sup>3 5</sup>

Olga Bolaños Solano<sup>4 5</sup>

No presente trabalho se utilizou-se o Modelo de Resposta Aleatorizada proposto por Odumade & Singh (2009), na pesquisa feita na Faculdade de Ciências Matemáticas (FCM), para estudar o comportamento de estudantes frente a questões sensíveis, por exemplo, o consumo de bebidas alcoólicas até o estado de ebriedad. A população estudada compreende os alunos matriculados no Semestre 2016-I pertencendo às quatro Escolas Profissionais da Faculdade de Ciências Matemáticas da Universidad Nacional Mayor de San Marcos no Peru. O tamanho da amostra foi calculado utilizando amostragem estratificada aleatória com o limite do erro de estimativa de 0,05 e um nível de confiança de 95% (Scheaffer, Mendenhall e Ott, 2001), e a informação fornecida pela Diretoria Acadêmica da FCM, dos estudantes de graduação matriculados no primeiro semestre do ano lectivo do 2016, o tamanho da amostra foi de 306 alunos, distribuídos proporcionalmente às quatro Escolas Profissionais consideradas.

**Palavras-chave:** *Perguntas sensíveis, Modelo de Resposta Aleatorizada, Tamanho de amostra.*

---

<sup>1</sup> FCM – UNMSM. e-mail: osolanod@unmsm.edu.pe

<sup>2</sup> FCM – UNMSM. e-mail: jernauh@gmail.com

<sup>3</sup> FCM – UNMSM. e-mail: fbartolog@unmsm.edu.pe

<sup>4</sup> FPS – URP. e-mail: olga\_bs\_89@hotmail.com

<sup>5</sup> Agradecimento ao Vice-Reitor de Pesquisa da UNMSM pelo apoio financeiro.

## Utilizando séries temporais na previsão do número de passageiros nos voos Brasileiros

Josiane Correia de Souza Carvalho <sup>1 3 4</sup>

Gislene Araújo Pereira <sup>2 3</sup>

A previsão do número de passageiros nos voos brasileiros é importante a fim de apontar ao setor a necessidade de melhorias. A variável utilizada foi passageiros pagos, que refere-se aos passageiros que ocupam assentos comercializados ao público de empresas de transporte aéreo de todo Brasil. Tais informações são mensais e retroativas até o ano de 2008 e foram disponibilizadas pela Agência Nacional de Aviação Civil- ANAC. O objetivo deste trabalho é indicar a movimentação futura do setor aéreo durante o ano de 2018. Pelo fato dos dados possuírem sazonalidade, devido a característica própria da variável, o modelo de séries temporais mais adequado é o modelo SARIMA. Foi utilizada série logaritmizada, e a nível de 5% de significância, pelo Teste de Cox Stuart e pelo Teste de Fisher, tal série não foi considerada estacionária em tendência nem em sazonalidade, o que foi corrigido após a primeira diferença e primeira diferença sazonal. Com a série estacionária em tendência e sazonalidade, foram ajustados os modelos obtidos com auxílio do gráfico de autocorrelação e autocorrelação parcial. Utilizando o critério de AIC foi escolhido o modelo SARIMA (0,1,0)(1,1,2). Os resultados da previsão demonstram que os meses em que espera-se maior número de passageiros em voos brasileiros são janeiro, julho e outubro de 2018, com números cerca de 4% superiores aos respectivos meses em 2017. O que indica ao setor aéreo a necessidade de melhorias frente ao aumento da demanda.

**Palavras-chave:** *Demanda do setor aéreo, Modelo SARIMA, Número de passageiros.*

---

<sup>1</sup>Programa de Pós Graduação em Estatística Aplicada e Biometria- UNIFAL.  
e-mail: [josianecorreia.carvalho@gmail.com](mailto:josianecorreia.carvalho@gmail.com)

<sup>4</sup>Agradecimento à CAPES pela bolsa.

<sup>2</sup>Instituto de Ciências Sociais Aplicadas- UNIFAL.

<sup>3</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## Validação psicométrica do Career Transitions Inventory (CTI), para profissionais brasileiros, por meio da Modelagem de Equações Estruturais

Naiara Caroline Ap. dos Santos <sup>1 3</sup>

Eraldo Schunk Silva <sup>2 3</sup>

Marlize Paulo da Silva <sup>2 3</sup>

O Modelo de Equações Estruturais (MEE), conhecido também por modelo LISREL, é um modelo linear que estabelece relações entre variáveis latentes e variáveis manifestas sob estudo. O termo Modelo de Equações Estruturais não designa apenas uma única técnica estatística, mas uma família de procedimentos relacionados, conhecida por diversos nomes, entre eles a análise de caminhos, a análise de estrutura de covariância, a análise de variáveis latentes e a análise fatorial confirmatória. É uma técnica de amplo uso e considera-se que sua função principal é a especificação e a estimação de modelos relações lineares entre variáveis. Assim, propõe-se aplicar a MEE para validar o questionário *Career Transitions Inventory* desenvolvido por Heppner et al. (1994), o qual identifica recursos psicológicos que os indivíduos possuem diante das transições de carreira. Ele possui 40 itens distribuídos em 5 fatores: prontidão, confiança, apoio, decisão independente e controle. A coleta de dados ocorreu via plataforma *online*, por meio de acessibilidade específica e efeito bola de neve. A amostra é composta por 533 profissionais com nível superior de escolaridade e em processo de transição de carreira.

**Palavras-chave:** *Modelos de Equações Estruturais, Análise Fatorial Confirmatória, Validação Psicométrica, Transição de Carreiras.*

---

<sup>1</sup>PBE - UEM. e-mail: [naicaroline2@gmail.com](mailto:naicaroline2@gmail.com)

<sup>2</sup>PBE - UEM.

<sup>3</sup>Agradecimento ao CNPq pelo apoio financeiro.

## Vida escolar: Agressor ou vítima? O uso da análise de correspondência simples na ocorrência de bullying.

Victor Ferreira da Silva <sup>1 3</sup>

Kelly Pereira de Lima <sup>2 3</sup>

Édipo Menezes da Silva <sup>2 3</sup>

Felipe Augusto Fernandes <sup>2 3</sup>

João Domingos Scalón <sup>2 3</sup>

O bullying é uma prática de atitudes agressivas, de maneira intencional e repetitiva, no ambiente escolar. Os maiores autores dessa prática são os alunos. O objetivo deste trabalho é identificar, mediante a ocorrência de bullying, se agredidos tendem a se tornarem agressores entre alunos do nono ano de uma escola pública estadual da cidade de Mariana/ MG. Os dados foram obtidos através da aplicação de um questionário que possui estruturas de dependências. Foram escolhidas duas perguntas: uma se o aluno foi agredido e a outra se já agrediu algum colega. Com relação aos itens, tem-se as seguintes alternativas para as duas perguntas: (1) descaracteriza a ocorrência; (2) caracteriza ser vítima/agressor indiretamente; (3) caracteriza ser vítima/agressor em potencial e (4) caracteriza ser praticante ou vítima. Para analisar os dados de forma adequada, levou-se em consideração essas estruturas e utilizou a técnica multivariada análise de correspondência simples. Esta técnica foi aplicada nos dados categóricos que levam em consideração as associações entre objetos caracterizados por atribuídos em uma matriz de dupla entradas. Além disso, esta técnica permite uma melhor interpretação dos dados, pois pode-se obter um gráfico *biplot*. Os resultados obtidos mostram que para as duas questões, a maior frequência se concentra na alternativa (1), ou seja, os alunos que responderam o questionário tendem a descaracterizar a ocorrência de bullying. Os resultados também mostraram que quem nunca sofreu ou praticou bullying, tende a, pelo menos, uma vez praticá-lo.

**Palavras-chave:** *Violência Escolar, Bullying, Análise Multivariada.*

---

<sup>1</sup>DES - UFLA. e-mail: [victor.est.ufla@gmail.com](mailto:victor.est.ufla@gmail.com)

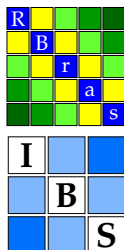
<sup>2</sup>DES - UFLA.

<sup>3</sup>Agradecimento ao CNPq, FAPEMIG e CAPES pelo apoio financeiro.

Realização



Departamento  
de Estatística



Apoio

