

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI



**ANAIS DO XIII ENCONTRO  
MINEIRO DE ESTATÍSTICA**

XIII ENCONTRO MINEIRO  
DE ESTATÍSTICA  
**MGEST**

02 E 03 DE OUTUBRO DE 2014

**Paulo César de Resende Andrade**  
**(Presidente da Comissão Organizadora)**  
**Marcel da Silva Toledo**  
**(Presidente da Comissão Científica)**

**ANAIS DO XIII ENCONTRO  
MINEIRO DE ESTATÍSTICA**

**XIII** ENCONTRO MINEIRO  
DE ESTATÍSTICA  
**MGEST**

**DIAMANTINA - MG**  
**OUTUBRO - 2014**

Ficha Catalográfica – Serviço de Bibliotecas/UFVJM  
Bibliotecário Anderson César de Oliveira Silva, CRB6 – 2618.

E56

Encontro Mineiro de Estatística (13. : 2014 : Diamantina, MG)  
Anais do XIII Encontro Mineiro de Estatística. – Diamantina:  
UFVJM, 2014.  
296 p. : il.

Presidente da Comissão Organizadora, Paulo César de Resende  
Andrade. Presidente da Comissão Científica, Marcel da Silva Toledo

1. Ciências exatas e da terra. 2. Probabilidade e Estatística.  
I. Título. II. Universidade Federal dos Vales do Jequitinhonha e  
Mucuri.

**CDD 519**

Elaborado com os dados fornecidos pelo(a) autor(a).

Apoio:



Realização:



**Coordenação Geral:**

Glaura da Conceição Franco – (UFMG)

**Comissão Organizadora Local:**

Paulo César de Resende Andrade – (ICT/UFVJM)  
(Presidente)

Emerson Cotta Bodevan – (FACET/UFVJM)

Geovane da Conceição Máximo – (FIH/UFVJM)

Josiane Magalhães Teixeira – (FACET/UFVJM)

Lucas Franco Ferreira – (ICT/UFVJM)

Marcelino Serretti Leonel – (ICT/UFVJM)

**Comissão Científica:**

Marcel de Toledo Vieira - (UFJF) (Presidente)

Antônio Policarpo Souza Carneiro – (UFV)

Eric Batista Ferreira - (UNIFAL)

Fernando Luiz Pereira de Oliveira - (UFOP)

João Domingos Scalon - (UFLA)

Marcos Santos de Oliveira - (UFSJ)

Nilson Luiz Castelucio Brito - (UNIMONTES)

Lourdes Coral Contreras Montenegro - (UFMG)

Priscila Neves Faria - (UFU)

Ricardo Luis dos Reis - (ICT/UFVJM)

# Habilidades Sensoriais de Grupos Heterogêneos de Consumidores de Cafés Especiais Discriminadas pelo Método MFACT

Paulo Cesar Ossani<sup>1</sup>

Marcelo Angelo Cirillo<sup>2</sup>

## 1 Introdução

A atual realidade dos mercados nacional e internacional de café aponta para a crescente demanda por cafés especiais, de sabor e aroma excepcionais e com características marcantes na doçura, acidez e corpo, e por produtos cujas qualidades ou características estejam intimamente relacionadas ao meio geográfico. Nesse cenário, o setor produtivo e o mercado estabelecem uma relação comercial transparente pautada na livre escolha por parte do consumidor e na agregação de valores e diferenciação de preços em função da qualidade do produto.

Neste cenário promissor em relação aos valores que os cafés especiais tendem a agregar no mercado, necessariamente, a qualidade dos cafés avaliada por meio da análise sensorial, considerando provadores treinados ou consumidores. A questão é que na situação em que envolve provadores treinados com a habilidade de discriminar pequenas diferenças entre as amostras, dada a homogeneidade entre os resultados, proporcionados pelas avaliações dos treinadores, os resultados obtidos por meio de uma metodologia estatística são precisos e coerentes.

Com o propósito de avaliar a concordância entre as respostas de diferentes grupos de consumidores com diferentes habilidades sensoriais, um experimento relacionado à análise sensorial de cafés especiais foi planejado consubstanciado pelo CEP. Em função dos resultados obtidos, como alternativa utilizou-se o método MFACT - *Multiple Factor Analysis for Contingency Tables* na avaliação da qualidade sensorial de cafés, sendo estes definidos por (A) Bourbon Amarelo, cultivado a altitudes maiores do que 1200 m.; (D) Ídem a (A) diferenciando apenas no preparo das amostras; (B) Acácia cultivados a altitude menores do que 1.100m; (C) Ídem ao (B) porém diferenciando o preparo das amostras.

---

<sup>1</sup> Mestrando em Est. Exp. Agrop., Depto Ciências Exatas, UFLA, e-mail: ossanipc@hotmail.com

<sup>2</sup> Prof. Adjunto IV, Depto Ciências Exatas, UFLA, e-mail: macufla@dex.ufla.br

## 2 Objetivo

Avaliar a concordância entre as respostas de diferentes grupos de consumidores com habilidades sensoriais diferentes, na análise de cafés especiais.

## 3 Metodologia

Para os propósitos de nosso estudo foram formados quatro grupos contendo respectivamente 52, 47, 32 e 43 indivíduos, sendo que o Grupo 1 foi formado por consumidores que receberam um treinamento sensorial; Grupo 2 e 3 por consumidores com diferentes níveis de experiência em análise sensorial de café e o Grupo 4 foi formado por profissionais na área de qualidade, produtividade e sensorial na pesquisa cafeeira. Para todos os indivíduos de cada grupo foram apresentados uma planilha, para que os mesmos dessem uma nota de 0 (zero) à 10 (dez), para as características: Nota Geral, Corpo, Acidez e Doçura para os cafés Bourbon Amarelo e Acáia com as características apresentadas na introdução.

Análise de Múltiplos Fatores para Tabelas de Contingências (MFACT - Multiple Factor Analysis for Contingency Tables) é uma técnica multivariada que compara as estruturas de várias tabelas de contingências com a mesma linha e colunas diferentes (ou vice-versa), geralmente construídas de diferentes amostras ou populações. Uma apresentação teórica detalhada da metodologia pode ser encontrada em Bécue-Bertaut e Pagès (2004).

Por estrutura de uma tabela, queremos dizer a relação entre linhas e colunas conforme expresso na análise de correspondência (AC) através de interfaces gráficas. As comparações têm de lidar com ambas as linhas e colunas, isto é, a estrutura induzida pelas linhas ao longo dos diferentes conjuntos de colunas, e também a estrutura induzida sobre as linhas pelos diferentes conjuntos de colunas, veja (BÉCUE-BERTAUT; PAGÈS, 2008).

O método consiste em tomar uma matriz  $Z$ , constituída da concatenação das subtabelas  $X_i$  balanceadas, em seguida fazendo;  $Q = \text{diag}(\sqrt{M}) \times Z \times \text{diag}(\sqrt{W})$ , e usando a Decomposição Valor Singular Generalizada sabemos que  $Q = \hat{U}\hat{\Delta}\hat{V}^T$  (ABDI, 2007) com  $\hat{U}^T M \hat{U} = \hat{V}^T W \hat{V} = I$ , onde  $U$  e  $V$  são as matrizes de autovetores e  $\Delta$  é a matriz de valores singulares, lembrando que  $M$  e  $W$  são as restrições das linhas e colunas respectivamente. Com isso obtemos os escores dos fatores globais que são dados por  $F = \hat{U}\Delta$ . Ainda temos que os autovalores são dados por  $\Delta_i^2$ . Com tudo isso, podemos calcular os escores dos  $i$  grupos de variáveis, pela equação;  $F_i = T \times X_i \times \text{diag}(W_i) \times V_i$ , onde  $T$  é número de subtabelas. Segue também que as inércias

entre os grupos são dadas por  $G = V \times \text{diag}(\sqrt{\Delta})$  e  $H = (G_{ij})^2 \times \text{diag}\left(\frac{1}{\Delta}\right)$ . Finalmente a equação  $P = w_i \times H_i$  nos leva às inércias procuradas, onde  $\lambda_s \times \sum_j^k P_{j,k}$  nos proporciona as inércias de cada Grupo, sendo k o número de variáveis de cada grupo e  $\lambda_s, s = 1, \dots, n$ , são os valores singulares de Q.

## 4 Resultados e discussões

Projetando o conjunto de dados de cada Café sobre a análise global das faixas dos indivíduos (Figura 1), teremos uma interpretação de como cada Café “interpreta” o espaço das faixas dos indivíduos. Note que na Figura 1 a posição de cada faixa de indivíduos na análise global é o baricentro (isto é, centroide) de suas posições para os Cafés. Para facilitar a interpretação, puxamos linhas que unem os Cafés com a posição global das faixas dos indivíduos.

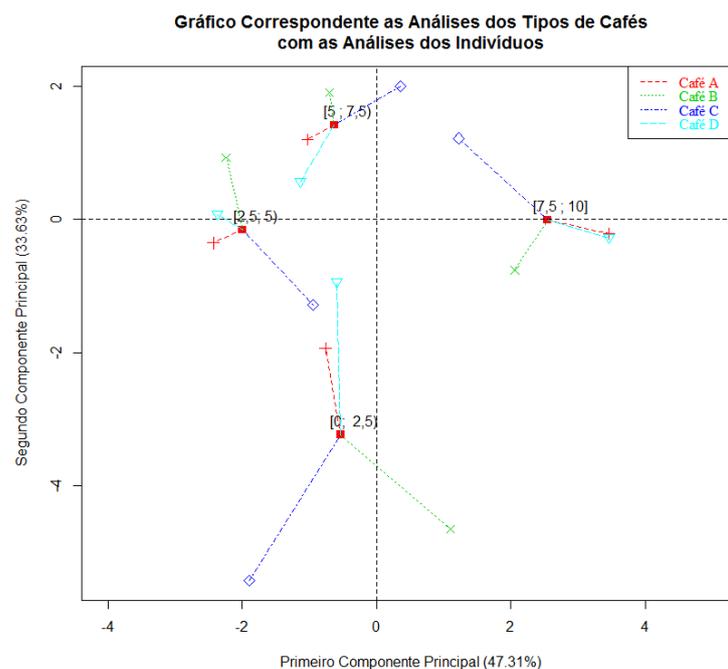


Figura 1 – Gráfico das Análises dos Tipos de Cafés com as Análises dos Indivíduos

Para fins de exemplificação tomemos a faixa de indivíduos [5; 7,5], para análise verifica-se que como estamos trabalhando com um centroide, podemos interpretar que os Cafés A e D possuem quase que a mesma órbita na faixa citada e que os Cafés B e C possuem órbitas distintas, observe que este padrão se repete para a maioria das outras faixas. Isto sugere que os Cafés B e C estão particularmente em discrepância com os Cafés A e D.

Agora temos que as relações entre os Cafés e a solução global são analisadas calculando a inércia parcial de cada Café por cada dimensão da análise global, o que segue na Tabela 1, corroborando com as afirmações relacionadas à similaridade dos Cafés A e D.

Tabela 1 - Valores referentes às inércias computadas nos três primeiros eixos para cada grupo de Tipos Cafés.

Grupos de Tipos Especiais de Cafés					
Eixo	A	B	C	D	Inércia total
1	0,9923	0,6568	0,3521	0,9881	2,9893
2	0,3633	0,7379	0,8532	0,1709	2,1254
3	0,2963	0,5101	0,2707	0,1272	1,2043

A partir das inércias obtidas em cada Café (Tabela 1), visando uma melhor interpretação, o gráfico das inércias (Figura 2) é gerado. Em síntese, nota-se uma concordância entre a discriminação dos Cafés A e D, com maior inércia (Tabela 1), em relação aos Cafés B e C, sendo do genótipo Acaia produzido em uma altitude inferior.

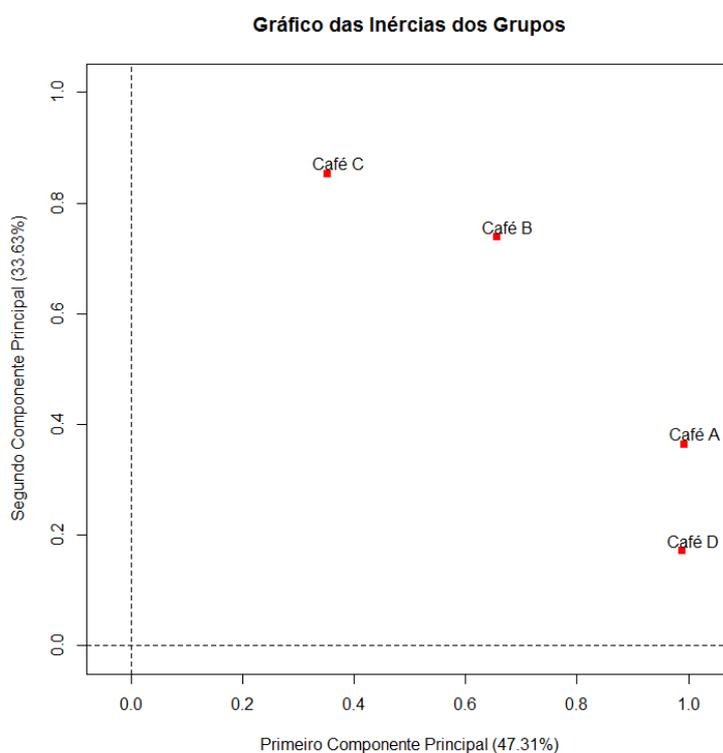


Figura 2 – Inércias dos Grupos de Cafés

## 5 Conclusões

Conclui-se que o método MFACT foi adequado na diferenciação das avaliações sensoriais realizadas por grupos distintos, de modo que, os resultados mais promissores foram observados nas similaridades dos Grupos dos Cafés A e D. Sugerindo que a análise sensorial do café Bourbon Amarelo não é refletida pela formação das pessoas.

## Referências

- [1] ABDI, H; Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD). In: Salkind NJ, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage; 2007.
- [2] BÉCUE-BERTAUT, M.; PAGÈS, J.; A principal axes method for comparing contingency tables: MFACT. *Computational Statistics & Data Analysis*, v. 45, p. 481-503, 2004.
- [3] BÉCUE-BERTAUT, M.; PAGÈS, J.; Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Computational Statistics & Data Analysis*, v. 52, p. 3255-3268, 2008.
- [4] ESCOFIER, B.; Analyse factorielle en référence à un modèle: application à l'analyse d'un tableau d'échanges. *Revue de Statistique Appliquée*, vol. 32, n° 4, p. 25-36, 1984.
- [5] GREENACRE, M.; BLASIUS, J.; *Multiple Correspondence Analysis and Related Methods*. 1th. ed. New York: Taylor and Francis Group, 2006.

# Inferência Bayesiana para uma extensão assimétrica do modelo de Grubbs

Fábio Rocha da Silva - DEst, UFMG<sup>1</sup>  
Lourdes C. Montenegro - DEst, UFMG

**Resumo:** O modelo de Grubbs é utilizado para comparar vários instrumentos de medição, e é comum supor que os termos aleatórios seguem uma distribuição normal. Neste trabalho estudamos aspectos de inferência bayesiana para o modelo de Grubbs, onde o valor não observado da covariável (variável latente) segue uma distribuição normal assimétrica (Montenegro et al. (2010)). As estimativas *a posteriori* dos parâmetros do modelo em estudo são calculados através do amostrador de *Gibbs*. Os resultados e métodos desenvolvidos neste trabalho são ilustrados através de um conjunto de dados.

**Palavras-chave:** modelo de Grubbs, distribuição normal assimétrica, amostrador de Gibbs.

## 1 Introdução

Regressão linear é uma das ferramentas estatísticas mais amplamente utilizados e tem sido objeto de um extenso estudo na literatura por mais de um século. Em várias aplicações a covariável não é diretamente observado (variável latente), por esta razão, esta covariável é medida com erros. Há uma extensa quantidade de trabalhos sobre este modelos, entre as principais referências podemos citar Fuller (1987). Os modelos com erros de medição são utilizados em modelos de regressão linear e modelos de regressão não-linear, modelos de análise fatorial, modelos com estruturas latentes e modelos de equações simultâneas. Estes modelos são também amplamente utilizados para comparar instrumentos de medida (ver, Barnett (1969); Theobald (1978); Shyr (1986); Bolfarine (1995)), que se diferenciam pelo custo, tempo gasto para medir e, outras características, tais como a eficiência. Uma classe particular dos modelos utilizados para comparar instrumentos de medida foi introduzido por Grubbs (1948) e Grubbs (1973). Este modelo é usualmente utilizado em estudos de comparação de métodos para avaliar a concordância relativa entre dois ou mais métodos analíticos (ou instrumentos), que medem a mesma quantidade de interesse na mesma escala. Seja  $x_i$  o valor resposta não observada (latente) correspondente a unidade  $i$  e  $y_{ij}$  o valor medido obtido com o instrumento  $j$  na unidade de  $i$ ,  $i = 1, \dots, n$  e  $j = 1, \dots, p$ , com  $p \geq 2$ . Supondo que o primeiro instrumento é a referência que serão comparados com os restantes  $p - 1$  instrumentos, o modelo de Grubbs normal é definida como

$$\mathbf{y}_i = \mathbf{a} + \mathbf{1}_p x_i + \boldsymbol{\epsilon}_i, \quad (1)$$

onde  $i = 1, \dots, n$ ,  $\mathbf{a} = (0, \alpha)^T = (0, \alpha_2, \dots, \alpha_p)^T$  e  $\mathbf{1}_p = (1, \dots, 1)^T$  são vetores  $p \times 1$ ;  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})^T$  e  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{ip})^T$  (o vetor erro), são vetores aleatórios  $p \times 1$  independente com  $\boldsymbol{\epsilon}_i \stackrel{\text{iid}}{\sim} N_P(0, D(\boldsymbol{\phi}))$  e  $x_i \stackrel{\text{iid}}{\sim} N_1(\mu_x, \phi_x)$ , em que  $D(\boldsymbol{\phi})$  denota a matriz diagonal com  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^T$ . Inferência para este modelo pode ser encontrado em Grubbs (1948) e Grubbs (1973).

---

<sup>1</sup>Contato: [fabiorochadasilva@ufmg.br](mailto:fabiorochadasilva@ufmg.br)

Embora a suposição de normalidade (ou simetria) é adequada, em muitas situações não é apropriado quando os dados apresentam comportamentos que diferem do modelo normal, como por exemplo, assimetria. O conjunto de dados estudado por Barnett (1969) apresenta esta característica e parece requerer transformação de modo a ser melhor aproximada pela distribuição normal. Portanto, usaremos uma classe geral de modelos que compreende a normal como um caso particular desta classe. A família normal assimétrica foi estudada por Azzalini (1985) em um contexto univariado. Nesta família de distribuições, um parâmetro de forma que regula a assimetria da distribuição, permite uma forma de transição contínua da não normalidade para normalidade.

Dizemos que o vetor aleatório  $y$  de dimensão  $k$  tem distribuição normal assimétrica multivariada com parâmetros de locação  $\mu$ , matriz escala  $\Psi$  (Positiva Definida) e parâmetros de assimetria  $\lambda$ , o qual será denotada por  $SN(\mu, \Psi, \lambda)$ , se sua função de densidade de probabilidade é dada por

$$f(y) = 2\phi_k(y|\mu, \Psi)\Phi_1(\lambda^T\Psi^{-1/2}(y - \mu)), \quad y \in \mathcal{R}, \quad (2)$$

onde  $\phi_k(y|\mu, \Psi)$  representa a função de densidade da distribuição normal  $k$ -variada com  $\mu$  o vetor de médias e  $\Psi$  a matriz de covariância,  $\Phi_1$  representa a função de distribuição acumulada da distribuição normal padrão, e  $\Psi^{-1/2}$  satisfaz  $\Psi^{-1/2}\Psi^{-1/2}\Psi^{-1}$ . Quando  $\lambda = 0$ , temos que  $y \sim N_k(\mu, \Psi)$ . A representação estocástica de uma variável aleatória normal de inclinação, o qual pode ser usado para simular realizações aleatórias a partir de  $y$ , é dada pela:

$$y \stackrel{d}{=} \mu + \Psi^{1/2}(\delta|T_0| + (\mathbb{I}_k - \delta\delta)^{1/2}T_1), \quad \text{com } \delta = \frac{\lambda}{\sqrt{1 + \lambda^T\lambda}}, \quad (3)$$

onde  $T_0 \sim N_1(0, 1)$  e  $T_1 \sim N_k(0, \mathbb{I}_k)$  são independentes, e “ $\stackrel{d}{=}$ ” significa “distribuída como”. Para mais detalhes sobre essa abordagem, consulte Arellano-Valle and Genton (2005). Note que em (3) quando  $k = 1$ , obtemos a distribuição normal univariada.

Este trabalho está organizado da seguinte forma. A Seção 2 apresenta a formulação do modelo de Grubbs normal assimétrico (MGNA). Na Seção 3, apresentamos um procedimento de inferência bayesiana para o modelo Grubbs normal assimétrico. Finalmente, um exemplo ilustrativo previamente analisada na literatura será novamente analisado na Seção 4.

## 2 O Modelo

Neste trabalho, para uma estimativa mais realista dos parâmetros do modelo, especialmente em situações em que os dados apresenta assimetria, assumimos que  $(x_i, \epsilon_i)^T$  segue uma distribuição  $(p + 1)$ -variada normal assimétrica, isto é,

$$\begin{pmatrix} x_i \\ \epsilon_i \end{pmatrix} \stackrel{iid}{\sim} SN_{p+1} \left( \begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, D(\phi_x, \phi), \begin{pmatrix} \lambda_x \\ 0 \end{pmatrix} \right) \quad (4)$$

onde  $i = 1, \dots, n$ ,  $(\mu_x, 0)^T$  é o vector de locação,  $D(\phi_x, \phi)$  representa a matriz de escala, e  $(\lambda_x, 0)^T$  é o parâmetro de assimetria, com  $D(\phi_x, \phi)$  denota uma matriz diagonal com elementos  $\phi_x$  e  $\phi$ . Do Lema 2.2 em Gupta e Chen (2004), pode-se mostrar que  $\epsilon_i$  e  $x_i$  são independentes com

$$\epsilon_i \stackrel{iid}{\sim} N_p(0, D(\phi)) \text{ e } x_i \stackrel{iid}{\sim} SN_1(\mu_x, \phi_x, \lambda_x), \quad (5)$$

onde  $i = 1, \dots, n$ . É claro que  $\phi_j$  está relacionada com a variação da medição feita pelo instrumento de  $j$ ,  $j = 1, \dots, p$ , e  $\phi_x$  para o parâmetro de escala da resposta latente, que tem variância  $\phi_x\{1 - 2\lambda^2/(\pi(1 + \lambda^2))\}$ . Este modelo, por exemplo, representa o conjunto de dados estudados por Barnett (1969), o qual mostra que a distribuição da capacidade vital do pulmão

humano não é simetricamente distribuídas na população e uma vez que  $\epsilon_i$  está relacionada com erro de medição, deve ser normalmente distribuído (com média zero). Além disso, estamos supondo que a variável latente ( $x$ ) segue uma distribuição normal assimétrica.

A partir da representação estocástica marginal de um vector aleatório normal assimétrico dado em (3), com  $t = |T_0|$ , o modelo definido em (4) pode ser escrita hierarquicamente como

$$\mathbf{y}_i|x_i \stackrel{ind}{\sim} N_p(\mathbf{a} + \mathbf{1}_p x_i, D(\phi)), \quad (6)$$

$$x_i|T_i = t_i \stackrel{ind}{\sim} N_1(\mu_x + \phi_x^{1/2} \times \delta_x t_i, \phi_x(1 - \delta_x^2)), \quad (7)$$

$$T_i \stackrel{iid}{\sim} HN_1(0, 1), \quad (8)$$

$i = 1, \dots, n$ , todos independentes, onde  $HN_1(0, 1)$  denota a distribuição half-normal univariada padrão e  $\delta_x = \lambda_x/(1 + \lambda_x^2)^{1/2}$ .

### 3 Inferência Bayesiana para o Modelo de Grubbs normal assimétrico(MGNA)

Em esta seção, implementamos a metodologia Bayesiana para o MGNA utilizando técnicas de MCMC (Markov Chain Monte Carlo), mais precisamente conhecido como amostrador de Gibbs. Para a implementação do algoritmo de Gibbs é necessário que seja realizado um estudo de sensibilidade sobre os desempenhos das distribuições *a priori* que serão utilizadas na estimação do MGNA. Bayes (2007); Rodriguez (2005) conduziram um estudo semelhante para modelos de regressão, neste estudo os autores sugeriram o uso da distribuição *a priori* para  $\lambda$  definida por:

$$\pi(\lambda) \propto \left(1 + \frac{\lambda^2}{k\sigma_t^2}\right)^{-\frac{k+1}{2}}, \quad (9)$$

onde,  $\sigma_t^2$  e  $k$  são os parâmetros de escala e os graus de liberdade de uma distribuição t-Student. Neste trabalho *a priori* (9) foi reescrita em forma hierárquica, apresentando a seguinte expressão:

$$\begin{aligned} \lambda|w &\sim N(0; \sigma_t^2) \\ w &\sim \text{Gama}\left(\frac{k}{2}; \frac{k}{2}\right). \end{aligned} \quad (10)$$

Este resultado segue se consideramos que  $\delta$  segue uma distribuição uniforme ( $U[-1; 1]$ ) e através de transformações, podemos observar diretamente, que  $\lambda$  segue uma distribuição t-Student, denotada por,  $\lambda \sim t(0; 1/2; 2)$ . Além disso, podemos utilizar uma *a priori* de Jeffreys para  $\lambda$ , que foi apresentada em LISEO (2006) para modelos de locação-escala normal assimétrico. Neste caso, *a priori* de Jeffreys,  $\pi_J(\lambda)$ , é dada por

$$\pi^J(\lambda) \propto \sqrt{\int_{-\infty}^{\infty} 2z^2 \phi(z) \frac{\pi(\lambda z)}{\Phi(\lambda z)} dz} \quad (11)$$

CHAIBUB NETO (2003.) desenvolveram após trabalhosas manipulações algébricas uma aproximação para a quantidade  $\lambda$  que aparece designada por  $\pi_J(\lambda)$ . Pode se mostrar que esta aproximação segue a distribuição *a priori* especificada por  $\pi_J(\lambda) \sim t(0; \pi^2/4; 1/2)$ , como foi mostrado em Rodriguez (2005).

Em seguida vamos implementar a metodologia Bayesiana para o MGNA usando a representação (6)-(8), com with  $\lambda = \varsigma/\sigma_*$  e  $\phi_x = \varsigma^2 + \sigma_*^2$ , segue-se a função de verossimilhança

completa associada com  $(Y, X, T)$  :

$$L(\theta|Y, X, T) = \prod_{i=1}^n \frac{1}{(2\pi)^{p/2} |D(\phi)|^{1/2}} \exp \left\{ \frac{-1}{2} (y_i - a - \mathbf{1}_p)' D(\phi)^{-1} (y_i - a - \mathbf{1}_p) \right\} \\ \times \frac{1}{\sqrt{(2\pi\sigma_*^2)}} \exp \left\{ \frac{-(x_i - \mu_x - \varsigma t_i)^2}{2\sigma_*^2} \right\} \times \sqrt{\frac{2}{\pi}} \exp \left\{ \frac{-t_i^2}{2} \right\}$$

Agora, para completar a especificação do modelo MGNA num contexto Bayesiano, presume-se que a distribuição a priori conjunta para  $\theta^* = (a, \phi, \mu_x, \varsigma, \sigma_*^2)$ , assumindo distribuições *a priori* independentes entre os parâmetros do modelo tem densidade dada por:

$$\pi(\theta^*) \propto \pi(a) \times \pi(\phi) \times \pi(\mu_x) \times \pi(\varsigma) \times \pi(\sigma_*^2) \quad (12)$$

onde,  $\alpha_i \sim N(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)$   $i = 2, \dots, p$ ,  $1/\phi_i \sim \text{Gama}(A_{\phi_i}, B_{\phi_i})$   $i = 1, \dots, p$ ,  $\mu_x \sim N(\mu_{xx}, \sigma_{xx}^2)$ ,  $1/\sigma_*^2 \sim \text{Gama}(A, B)$  e  $\varsigma \sim t(0; \sigma_t^2; g)$ , onde *a priori* para  $\varsigma$  pode ser obtida através da mistura  $b|w \sim N(0, \frac{\sigma_t^2}{w})$  e  $w \sim \text{Gama}(\frac{g}{2}; \frac{g}{2})$ .

Combinando a função de verossimilhança (12) e as distribuições a priori adotadas, a distribuição *a posteriori* conjunta de  $\theta^*$ . Pode-se notar que a esta distribuição *a posteriori* conjunta é intratável analiticamente porque a integração da densidade a posteriori não é fácil de ser feita. Uma alternativa, é considerar o algoritmo amostrador de Gibbs.

O amostrador de Gibbs funciona com a retirada de amostras de forma iterativa a partir das distribuições *a posteriori* condicionais derivadas dos parâmetros. Para o modelo MGNA temos o seguinte resultado:

**Proposição 1** *Assumindo distribuições a priori independentes entre os parâmetros do modelo de Grubbs, ou seja, fazendo, então, suposições de que  $\alpha_i \sim N(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)$   $i = 2, \dots, p$ ,  $1/\phi_i \sim \text{Gama}(A_{\phi_i}, B_{\phi_i})$   $i = 1, \dots, p$ ,  $\mu_x \sim N(\mu_{xx}, \sigma_{xx}^2)$ ,  $1/\sigma_*^2 \sim \text{Gama}(A, B)$  e  $\varsigma \sim t(0; \sigma_t^2; g)$ , onde a priori para  $\varsigma$  pode ser obtida através da mistura  $b|w \sim N(0, \frac{\sigma_t^2}{w})$  e  $w \sim \text{Gama}(\frac{g}{2}; \frac{g}{2})$ , teremos a distribuição a posteriori dada por*

$$\pi(\theta^* = (a, \phi, \mu_x, \varsigma, \sigma_*^2) | y_i) \propto \pi(\theta^*) \times L(\theta^*; z)$$

que conduz as seguintes distribuições a posteriori condicionais próprias:

$$a) \pi(\alpha_j | \sigma_*^2, \varsigma, \mu_x, w, \phi) \sim N \left( \frac{n\sigma_{\alpha_j}^2 \sum_{j=1}^n (y_{ij} - x_i) + \phi_j \mu_{\alpha_j}}{\phi_j + n\sigma_{\alpha_j}^2}; \frac{\phi_j \sigma_{\alpha_j}^2}{\phi_j + n\sigma_{\alpha_j}^2} \right), \quad j = 2, \dots, p$$

$$b) \pi(1/\sigma_*^2 | a, \varsigma, \mu_x, w, \phi) \sim \text{Gama} \left( \left( A + \frac{p}{2} \right); B + \frac{\sum_{i=1}^n (x_i - \mu_x - \varsigma t_i)^2}{2} \right)$$

$$c) \pi(\mu_x | a, \varsigma, \phi, w, \sigma_*^2) \sim N \left( \frac{\sigma_{xx}^2 \sum_{i=1}^n (x_i - \varsigma t_i) + \sigma_*^2 \mu_{xx}}{\sigma_*^2 + p\sigma_{xx}^2}; \frac{\sigma_*^2 \sigma_{xx}^2}{\sigma_*^2 + p\sigma_{xx}^2} \right)$$

$$d) \pi(1/\phi_1 | a, \varsigma, \mu_x, w, \sigma_*^2) \sim \text{Gama} \left( \left( A_{\phi_1} + \frac{n}{2} \right); B_{\phi_1} + \frac{\sum_{j=1}^n (y_{1j} - x_i)^2}{2} \right)$$

$$e) \pi(1/\phi_j|a, \varsigma, \mu_x, w, \sigma^{2*}) \sim Gama \left( \left( A_{\phi_j} + \frac{n}{2} \right); B_{\phi_j} + \frac{\sum_{j=1}^n (y_{ij} - \alpha_j - x_i)^2}{2} \right), \quad j = 2, \dots, p$$

$$f) \pi(\varsigma|a, \phi, \mu_x, w, \sigma^{2*}) \sim N \left( \frac{\frac{\sigma_t^2}{w} \sum_{i=1}^p [(x_i - \mu_x)t_i]}{\sigma_*^2 + \frac{\sigma_t^2}{w} \sum_{i=1}^p t_i^2}; \frac{\frac{\sigma_t^2}{w} \sigma_*^2}{\sigma_*^2 + \frac{\sigma_t^2}{w} \sum_{i=1}^p t_i^2} \right)$$

$$g) \pi(w|a, \phi, \mu_x, \varsigma, \sigma^{2*}) \sim Gama \left( \frac{g+1}{2}; \frac{4\sigma_t^2 + \varsigma^2 g}{2\varsigma^2} \right)$$

$$h) \pi(t_i|a, \phi, \mu_x, \varsigma, \sigma^{2*}, w) \sim Half-Normal \left( \frac{x_i - \mu_x}{\varsigma^2 + \sigma_*^2}; \frac{\sigma_*^2}{\varsigma^2 + \sigma_*^2} \right)$$

**Demonstração:** Depois de exaustiva manipulação algébrica, trabalhando diretamente com a função de verossimilhança completa. Obtemos a distribuição condicional de um particular parâmetro,  $\theta_i^*$ , condicionado em relação às outros parâmetros e deixando  $\theta_i^*$  livre.

Então, temos as condições necessárias para garantir a convergência do amostrador de Gibbs e este pode ser implementado da seguinte forma

1. . Atribua valores iniciais para o vetor  $\theta^{*(0)} = (a^{(0)}, \phi^{(0)}, \mu_x^{*(0)}, \varsigma^{(0)}, 1/\sigma_*^{2(0)})$
2. Gere  $w^{(j)}, t_i^{(j)}$ , utilizando os valores iniciais fixados;
3. Atualize  $\theta^{*(j)}$  através de suas respectivas distribuições;
4. Repita os passos 2 e 3 para  $j = 1, \dots, N$ ; e
5. Descarte os  $B$  primeiros valores ( $B \leq N$ ). .

Os  $N - B$  valores obtidos iterativamente serão utilizados para a estimação dos parâmetros de interesse.

## 4 Aplicação

Nesta seção, um conjunto de dados reais é utilizado para desenvolver a metodologia utilizada. Barnett (1969) estudaram dados sobre medições da capacidade vital do pulmão humano em um grupo comum de 72 pacientes, utilizando dois instrumentos (padrão e novo), explorados por operadores qualificados e não qualificados. Como visto em Bolfarine (1995), os quatro instrumentos foram comparados da seguinte forma

- Instrumento 1: instrumento padrão e operador qualificado;
- Instrumento 2: instrumento padrão e operador não qualificado;
- Instrumento 3: Novo instrumento e operador qualificado;
- Instrumento 4: Novo instrumento e operador não qualificado. Consideramos as medições dividido por 100, a fim de melhorar a estabilidade numérica.

Utilizando o amostrador de Gibbs para calcular as estimativas dos parâmetros do modelo. Foi utilizado o software *OpenBUGS* para a implementação do modelo em estudo, sendo que o *OpenBUGS* é utilizado para simular distribuições *a posteriori* e na atualidade é muito utilizado por pesquisadores que implementam algoritmos MCMC.

Desta forma foi ajustado um modelo através pelo *OpenBUGS* e que as distribuições *a priori* consideradas no Teorema 1, são especificadas por  $\alpha_i \sim N(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)$   $i = 2, \dots, p$ ,  $1/\phi_i \sim \text{Gama}(A_{\phi_i}, B_{\phi_i})$   $i = 1, \dots, p$ ,  $\mu_x \sim N(\mu_{xx}, \sigma_{xx}^2)$   $1/\sigma^2 \sim \text{Gama}(A, B)$  e  $\varsigma \sim t(0; \sigma_t^2; g)$ , com os hiperparâmetros, todos conhecidos (a exceção de  $g$ , que assumimos uma distribuição uniforme). A escolha destas distribuições serão baseadas nas especificações usuais da teoria de regressão linear. A seguir serão utilizados os seguintes valores para os hiperparâmetros:

$$\begin{aligned} \mu_{\alpha_i} &= \bar{y}_i - \mu_x x \\ \sigma_{\alpha_i}^2 &= \text{vary}_i \\ A_{\phi_i} &= B_{\phi_i} = 2 \\ \mu_{xx} &= 0.001 \\ \sigma_t^2 = \sigma_{xx}^2 &= 10000 \\ A &= B = 2 \end{aligned}$$

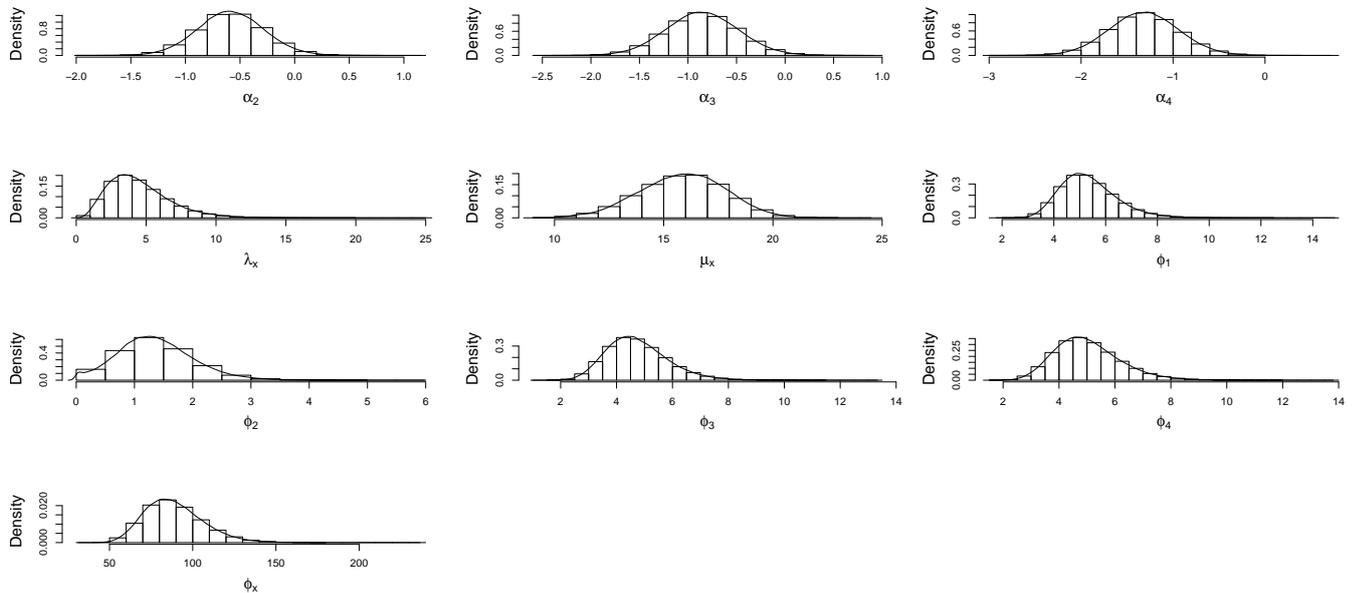
para esta implementação foram geradas 500000 iterações dentro do amostrador de Gibbs sendo descartadas as 5000 primeiras no processo de aquecimento das cadeias (*burn-in*). Os resultados desta implementação são apresentados juntamente com os resultados de Montenegro et al. (2010) na Tabela 4. A Figura 1 mostra os histogramas com a função de densidade aproximada via núcleo estimador (Scott et al.; 1977).

Tabela 1: Resultados da estimação do modelo de Grubbs normal assimétrico via Algoritmo EM em uma abordagem clássica (Montenegro et al.; 2010) e os resultados da estimação do modelo de Grubbs normal assimétrico via MCMC para o conjunto de dados de Barnett

Parâmetro	C		Bayesiano	
	Média	Desvio padrão	Média	Desvio padrão
$\alpha_2$	-0.7042	0.2961	-0.5951	0.2745
$\alpha_3$	-0.9750	0.3649	-0.8655	0.3046
$\alpha_4$	-1.4389	0.3693	-1.3291	0.3589
$\lambda_x$	5.6763	3.8284	4.4991	2.1886
$\mu_x$	12.1559	1.3918	13.4197	1.0151
$\phi_1$	5.0611	1.2888	5.3352	1.0711
$\phi_2$	1.2516	0.6640	1.3607	0.5242
$\phi_3$	4.5264	0.9997	4.7577	0.8984
$\phi_4$	4.7577	1.0646	5.0227	0.8154
$\phi_x$	168.91	39.9448	101.58	21.2468

De acordo com a Tabela 4 podemos observar que o valor do desvio padrão da maioria das estimativas dos parâmetros *a posteriori* no modelo de Grubbs Assimétrico são menores em relação as estimativas do modelo obtidas via algoritmo EM. Se compararmos o trabalho de Montenegro et al. (2010), as estimativas dos parâmetros de variabilidade obtiveram uma melhoria em sua precisão. Vale ressaltar que o valor zero não pertence ao intervalo de credibilidade de  $\lambda$ , o que evidencia que os dados possuem uma assimetria e, pelo intervalo de credibilidade, esta assimetria é positiva.

Figura 1: O Modelo *Grubbs* Normal Assimétrico Bayesiano: Densidades *a posteriori* aproximadas para  $\alpha_2$ ,  $\alpha_3$ ,  $\alpha_4$ ,  $\lambda_x$ ,  $\mu_x$ ,  $\phi_1$ ,  $\phi_2$ ,  $\phi_3$ ,  $\phi_4$ , e  $\phi_x$ .



## 5 Conclusão

Neste trabalho, propomos uma estratégia Bayesiana para estimação uma extensão do modelo de erro de medição *Grubbs* usando a distribuição normal assimétrica. A estimativa dos parâmetros é feita via um algoritmo MCMC, no caso o amostrado de Gibbs, implementado com expressões das condicionais completas dos parâmetros de interesse. Ao se comparar com o trabalho de Montenegro et al. (2010), a abordagem Bayesiana forneceu estimativas mais estáveis dos parâmetros de interesse

## Referências

- Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions, *Journal of Multivariate Analysis* **96**(1): 93–116.
- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian journal of statistics* pp. 171–178.
- Barnett, V. D. (1969). simultaneous pairwise linear structural relationships, *Biometrics* **25**: 129–142.
- Bayes, CL & Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution, *Brazilian Journal of Probability and Statistics* **21**(2): 141–163.
- Bolfarine, H. & Galea-Rojas, M. (1995). Maximum likelihood estimation of simultaneous pairwise linear structural relationships., *Biom J* **37**: 673–689.
- CHAIBUB NETO, E. & BRANCO, M. D. (2003.). Bayesian reference analysis for binomial calibration problem., *Technical report*, IME-USP.
- Fuller, W. (1987). *Measurement error models*, Wiley, New York.

- Grubbs, F. (1948). On estimating precision of measuring instruments and product variability, *Journal American Statistical Society* **43**: 243–264.
- Grubbs, F. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments, *Technometrics* **15**: 53–66.
- LISEO, B. & LOPERFIDO, N. (2006). A note on reference priors for the scalar skew-normal distribution., *Journal of Statistical Planning and Inference* **136**(2): 373–389.
- Montenegro, L. C., Lachos, V. H. and Bolfarine, H. (2010). Inference for a skew extension of the grubbs model, *Statistical Papers* **51**(3): 701–715.
- Rodriguez, C. L. B. (2005). *Inferência bayesiana no modelo normal assimétrico*, Master’s thesis, Universidade de São Paulo.
- Scott, D. W., Tapia, R. A. and Thompson, J. R. (1977). Kernel density estimation revisited, *Nonlinear Analysis: Theory, Methods & Applications* **1**(4): 339–372.
- Shyr, I. & Gleser, L. (1986). Inference about comparative precision in linear structural relationships., *Journal Statistical Planing Inference* **14**: 339–358.
- Theobald, C.M. & Mallison, J. (1978). Comparative calibration, linear structural relationship and congeneric measurements, *Biometrics* **34**: 35–45.

# Experimentos em quadrados latinos balanceados: uma apresentação algébrica e computacional

Tatiane Carvalho Alvarenga <sup>1</sup>

**Resumo:** Quadrados latinos balanceados são ferramentas muito utilizadas em estatística experimental, por exemplo, em ensaios rotacionais, em que os tratamentos são aplicados em uma mesma unidade experimental sequencialmente. A obtenção de quadrados latinos em geral segue regras algébricas de construção, a partir de operações básicas sobre o conjunto de números inteiros positivos, mais especificamente, sobre as classes de resto módulo ( $n$ ). A obtenção de quadrados latinos balanceados é mais complicada e exige uma teoria mais complexa, sendo o número de quadrados latinos balanceados para  $n$  maior ou igual a sete desconhecido. Neste trabalho utilizou-se a função *williams* do pacote *randomizeBE* do software *R*, na obtenção de quadrados latinos balanceados e de uma rotina para a estimação dos mesmos, considerando-se a ordem igual a  $n$ . Foram obtidos os números de quadrados latinos necessários na definição de um grupo balanceado, considerando-se até a ordem sete, sendo que a rotina mostrou-se bastante eficiente.

**Palavras-chave:** *DQL, Delineamento, Balanceamento, randomizeBE.*

## 1 Introdução

No delineamento em quadrado latino (DQL) cada tratamento deve aparecer uma única vez em cada linha e em cada coluna. Assim, no planejamento de tais ensaios podem ser utilizadas diferentes configurações na atribuição dos tratamentos às linhas e colunas. Nos quadrados latinos balanceados, cada tratamento deve ser precedido e seguido por cada um dos outros tratamentos o mesmo número de vezes. Em razão do elevado número de quadrados latinos necessários para garantir o balanceado nos casos de se ter maior número de tratamentos ( $n > 7$ ), o número exato conhecido de possíveis quadrados latinos é menor ou igual a sete. Para  $n$  maior ou igual a oito tem-se apenas o número aproximado. O número de quadrados latinos balanceados é desconhecido, apenas é apresentado em SENN (2006 citado por Wang; Gong, 2009) que para quatro tratamentos existem seis quadrados latinos balanceados. Em razão desse fato, uma simulação probabilística com o software *R*, foi desenvolvida para calcular o número total de quadrados latinos balanceados para  $n$  até sete em um tempo razoável e uma função do *R* que gera tais quadrados latinos balanceados.

## 2. Referencial Teórico

Um quadrado latino é dito estar na forma padrão se a primeira linha e coluna estão na ordem natural. A relação entre o número de quadrados latinos e o número de quadrados latinos na forma padrão é dada por:

**Teorema 2.1:** Para cada  $n \geq 2$  o número total  $L_n$  de quadrados latinos de ordem  $n$  é dada por:

$$L_n = n! (n - 1)! I_n, \quad (1)$$

em que  $I_n$  é o número de quadrados latinos de tamanho  $n$  na forma padrão.

O número total de quadrados latinos é desconhecido para  $n > 11$ . O melhor resultado possível é dado por (LINT; WILSON, 1992), ou seja,

**Teorema 2.2:** O número quadrados latinos de ordem  $n$  satisfaz (2):

$$\prod_{k=1}^n (k!)^{n/k} \geq L_n \geq \frac{(n!)^{2n}}{n^{n^2}} \quad (2)$$

Em que  $k = 1, \dots, n$ .

Este último teorema pode ser demonstrado utilizando-se conceitos sofisticados de análise combinatória. A tabela 1 apresenta o número de quadrados latinos até a ordem 11, de acordo com Laywine e Mullen (1998).

**Tabela 1:** Número de quadrados latinos de tamanho  $n$

$n$	Quadrados latinos na forma padrão	Todos os quadrados latinos
2	1	2
3	1	12
4	4	576
5	56	161280
6	9408	812851200
7	16942080	61479419904000
8	$> 5,35 \times 10^{11}$	$> 1,09 \times 10^{20}$
9	$> 3,77 \times 10^{17}$	$> 5,52 \times 10^{27}$
10	$> 7,5 \times 10^{24}$	$> 9,9 \times 10^{36}$
11	$> 5,3 \times 10^{33}$	$> 7,7 \times 10^{47}$

Nos quadrados latinos balanceados, cada símbolo, o qual representa um tratamento, deve ser precedido e seguido por cada um dos outros símbolos o mesmo número de vezes. Portanto, a ordem dos símbolos a serem atribuídos no quadrado latino não pode ser de forma totalmente aleatória. Portanto, não é qualquer forma casualizada de quadrado latino que garante a execução de um quadrado latino balanceado.

Para o caso de se ter um número par de símbolos, para que seja garantido o balanceamento, é necessário apenas um quadrado latino. Porém, quando esse número é ímpar, o balanceamento é atingido com dois quadrados latinos. Para número par de símbolos, a ideia básica é: os símbolos são indexados pelas classes de congruência módulo  $n$ , isto é pelos elementos do  $Z_n$ . A primeira linha deve ser de tal forma que as diferenças entre os termos consecutivos sejam todos os elementos de  $Z_n$ . Para  $n$  símbolos uma solução é a primeira linha ser dada por:

$$0 \quad 1 \quad n-1 \quad 2 \quad n-2 \quad 3 \quad n-3 \dots\dots\dots n/2 . \quad (3)$$

Na tabela 2, ilustra-se a construção de um quadrado latino balanceado com quatro símbolos (tratamentos), sendo que a primeira linha do quadrado é dada, conforme definição apresentada em (3), e as demais são dadas pela adição sequencial de uma unidade em módulo quatro.

**Tabela 2:** Balanceamento com quatro símbolos.

0	1	3	2	linha 1
1	2	0	3	(linha1+1)(mod4)
2	3	1	0	(linha2+1)(mod4)
3	0	2	1	(linha3+1)(mod4)

Quando o número de símbolos é ímpar, o balanceamento pode ser obtido com dois quadrados latinos de ordem  $(n \times n)$ . A ideia é que nas duas primeiras linhas as diferenças entre os termos consecutivos sejam todos os elementos de  $Z_n$  em que cada valor ocorra duas vezes. Para  $n$  símbolos uma solução é a primeira linha ser dada por:

$$0 \quad 1 \quad n-1 \quad 2 \quad n-2 \quad 3 \quad n-3 \dots\dots\dots(n+1)/2 . \quad (4)$$

Na definição da linha do segundo quadrado latino balanceado, utiliza-se o espelho da linha dada em (4), ou seja,

$$(n+1)/2 \dots\dots\dots n-3 \quad 3 \quad n-2 \quad 2 \quad n-1 \quad 1 \quad 0 . \quad (5)$$

O planejamento em quadrado latino balanceado pode ser facilmente realizado utilizando-se o software R (R Core Team, 2012). Utilizando-se a função *williams(ntmt, tmts)* do pacote *randomizeBE* (LABES, 2012), podendo-se obter configurações de quadrados latinos balanceados para qualquer número de tratamentos.

### 3. Material e Métodos

Neste trabalho foi utilizado o software R (R Core Team, 2012), mais especificamente a função *williams* do pacote *randomizeBE* (LABES, 2012), para ilustrar o planejamento de quadrado latino balanceado considerando diferentes números de tratamentos.

Como a definição do número de possíveis quadrados não pode ser demonstrada de forma trivial, foi desenvolvida uma rotina no software R para a determinação do número médio de quadrados latinos balanceados de uma ordem  $n$ .

### 4. Resultados e Discussão

Ao executar a função *williams* do pacote *randomizeBE* (LABES, 2012) no software R (R Core Team, 2012), tem-se um único quadrado latino balanceado quando o número de símbolos é par e dois quadrados latinos quando esse número é ímpar. Os quadrados são obtidos de forma aleatória, de todas as configurações possíveis.

Uma vez carregado o pacote *randomizeBE*, basta usar a função *williams(ntmt, tmts)* para obtenção do(s) quadrado(s) latino(s) balanceado(s). Os argumentos da função são *ntmt* e *tmts*, os quais referem-se, respectivamente, ao número de símbolos e a um vetor com a identificação dos símbolos. Caso não seja fornecido o vetor para *tmts*, o default é considerar uma sequência de letras maiúsculas A, B, C, .... A seguir são apresentados exemplos para diferentes números de símbolos.

Caso sejam considerados apenas quatro símbolos, basta digitar na linha de comando do R:

```
> williams(4)
```

obtendo-se como saída:

```
[1] "ADCB" "BCDA" "CABD" "DBAC".
```

No caso de quatro símbolos, existem seis possíveis quadrados latinos balanceados (SENN, 2006; citado por Wang e Gong, 2009). Para obtê-los no R, basta repetir o comando *williams(4) ..*

No caso de cinco símbolos tem-se:

```
> williams (5)
```

obtendo-se como saída:

```
[1] "ADEBC" "BCDEA" "CEBAD" "DBACE" "EACDB"  
[2] "AEDCB" "BDCAE" "CBEDA" "DABEC" "ECABD"
```

Na literatura foi encontrado o número de possíveis quadrados latinos balanceados apenas para configurações com no máximo quatro símbolos. A demonstração algébrica não é trivial. Assim, foi desenvolvido no software *R* (R Core Team, 2012) uma rotina para a determinação do número de quadrados latinos balanceados de uma ordem  $n$  de acordo com a função *williams* do pacote *randomizeBE* (LABES, 2012).

Executando a rotina para os casos de dois a sete tratamentos obtiveram-se os resultados apresentados na Tabela 3. Nessa tabela também consta o número de possíveis quadrados latinos e pode-se observar que a quantidade de quadrados latinos balanceados é bem menor comparado aos quadrados latinos.

Na utilização da rotina recomenda-se aumentar o valor de  $N$ , conforme aumenta-se o número de tratamentos.

**Tabela 3:** Número total de quadrados latinos e quadrados latinos balanceados obtidos pela função *williams*, do pacote *randomizeBE*.

$N$	Número total de quadrados latinos	Número de quadrados latinos balanceados
2	2	1
3	12	1
4	576	6
5	161280	12
6	812851200	120
7	61479419904000	360
8	$> 1,09 \times 10^{20}$	Não calculado
9	$> 5,52 \times 10^{27}$	Não calculado
10	$> 9,9 \times 10^{36}$	Não calculado
11	$> 7,7 \times 10^{47}$	Não calculado

## 5. Conclusão

A função *williams* do pacote *randomizeBE* (LABES, 2012) do software *R* (R Core Team, 2012) é simples e prática para se obter quadrados latinos balanceados. A rotina é eficiente na definição dos quadrados latinos balanceados até  $n = 7$ .

## 6. Referências

- [1] BOSE, R; B. MANVEL. **Introduction to Combinatorial Theory**. New Jersey: John Wiley & Sons, 1984.
- [2] LABES, D. **RandomizeBE: function to create a random list for crossover studies**. R package version 0.3-1. <http://CRAN.R-project.org/package=randomizeBE>. 2012.
- [3] LAYWINE, C., G. MULLEN. **Discrete Mathematics Using Latin Squares**. Canadá: John Wiley & Sons. 1998.
- [4] LINT, J. H., WILSON, R. M. **A Course in Combinatorics**. Cambridge University Press, 1992.
- [5] R Development Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>. 2012.
- [6] SANTOS, J. P. de O. **Introdução à teoria dos números**. 3. ed. Rio de Janeiro: IMPA, 2007. 198 p.
- [7] WANG, B.; WANG, X.; GONG, L. The Construction of a Williams Design and Randomization in Cross-Over Clinical Trials Using SAS. **Journal of Statistical Software**, vol 29. 2009.
- [8] WILLIAMS, E .J. Experimental designs for the estimation of residual effects of treatments. **Australian Journal of Scientific Research**, Series A: Physical Sciences, v.2, p.149-168, 1949.

# Aspectos inferenciais no Modelo de Crédito Parcial Generalizado da Teoria da Resposta ao Item

Eduardo Vargas Ferreira <sup>1 3</sup>

Caio Lucidius Naberezny Azevedo <sup>2 3</sup>

**Resumo:** O Modelo de Crédito Parcial Generalizado (MCPG) pertence à família dos modelos para as respostas politômicas graduais da Teoria de Resposta ao Item (TRI). É apropriado na modelagem de itens (questões) cujas respostas pertencem a alguma escala gradual. Embora ele seja de grande importância, não tem recebido a devida atenção, em particular, na literatura em Português. Mesmo na internacional, alguns pontos não têm sido estudados em detalhes. Com isto em mente, o presente projeto apresentou os principais aspectos inferenciais (sob o enfoque frequentista) do MCPG. Especificamente, avaliou-se graficamente como mudanças nos valores dos parâmetros influenciam o comportamento da curva característica do item (CCI). Estudou-se as condições necessárias para assegurar a identificabilidade do modelo. Tais discussões serviram como base para a consecução do principal objetivo: desenvolver e implementar computacionalmente os métodos de estimação frequentista do MCPG. Através de estudos de simulação, avaliou-se a acurácia das estimativas considerando diversas situações de interesse prático. Os resultados indicaram que todos os métodos recuperam bem os parâmetros. Além disso, verificou-se os fatores de maior impacto na precisão das estimativas dos traços latentes (número de itens e categorias), dos parâmetros de discriminação (número de indivíduos e categorias) e parâmetros de dificuldade (número de itens, categorias e indivíduos).

**Palavras-chave:** *Teoria da Resposta ao Item, Identificabilidade, Inferência frequentista.*

---

<sup>1</sup>IIMECC - UNICAMP: Universidade Estadual de Campinas. Email: [eduardozenus@yahoo.com.br](mailto:eduardozenus@yahoo.com.br)

<sup>2</sup>IIMECC - UNICAMP: Universidade Estadual de Campinas. Email: [cnaber@ime.unicamp.br](mailto:cnaber@ime.unicamp.br)

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

# 1 Introdução

Em muitos instrumentos de medida (como provas, questionários de avaliação de produtos etc.), os itens (questões) que os compõem apresentam alternativas que pertencem a alguma escala gradual. Por exemplo, num questionário de avaliação de um produto por parte de consumidores, as categorias de resposta poderiam variar de zero (ruim) a cinco (excelente), de modo discreto. Ou, numa avaliação cognitiva, como no vestibular da UNICAMP, questões dissertativas são apresentadas aos candidatos, e às suas respostas são atribuídos escores que varia de zero (erro total) à quatro (acerto total), de forma discreta. Nessas situações, tem-se interesse em avaliar os traços latentes (dos indivíduos) envolvidos. No primeiro, seria a avaliação do produto por parte dos consumidores, enquanto na segunda seria o conhecimento dos vestibulandos. Em ambos os casos, o Modelo de Crédito Parcial Generalizado (MCPG) poderia ser utilizado, veja [1] e [2]. O MCPG pertence à família de modelos para as respostas politômicas graduais da Teoria de Resposta ao Item (TRI). Embora ele seja de grande importância, não tem recebido a devida atenção na literatura, em particular, na literatura em Português. Mesmo na internacional, alguns pontos não têm sido estudados em detalhes como: interpretação dos parâmetros, estudo sobre a falta de identificabilidade e avaliação do desempenho de estimadores.

## 2 Objetivos

O objetivo do presente trabalho é estudar em maiores detalhes os principais aspectos inferenciais relacionados ao MCPG. Tais aspectos consistem em: interpretação dos parâmetros, particularmente aqueles relacionados aos itens, e as influências que os mesmos possuem entre si. Será feito, essencialmente, através de análises gráficas da função de resposta à categoria do item, veja [1], [3] e [4]. Aspectos ligados a identificabilidade, verificar-se-á as condições necessárias para assegurar tal característica, com relação a aplicar restrições tanto nos parâmetros dos itens quanto nos traços latentes. Devido à estrutura complexa do MCPG, é necessário restringir ambas as classes de parâmetros. E estudar de que forma isso se reflete na interpretação do modelo bem como o quanto afeta na estimação dos parâmetros de interesse. Que será desenvolvida e implementada computacionalmente (em linguagem  $\mathbf{R}$ ) via Máxima Verossimilhança (MV), Máxima Verossimilhança Perfilada (MVP) e Máxima Verossimilhança Marginal-Perfilada (MVMP). Além de um estudo de simulação visando avaliar o desempenho dos métodos supracitados, com relação a características importantes como: vício, variância e erro quadrático médio. Considerando diversas situações de interesse prático, formados pela escolha de diferentes números de indivíduos (NE), número de itens (NI) e categorias (NC).

### 3 Metodologia

A base das metodologias estudadas se encontra em [1], [2] e [4]. Tais trabalhos introduzem os problemas em questão e discutem, até certo ponto, alguns aspectos relacionados ao objeto de estudo. A interpretação dos parâmetros do MCPG é feita através da visualização (análise gráfica) das CCI's. Avalia-se como o aumento/diminuição dos parâmetros afetam o comportamento das curvas (probabilidades de escolha) para cada categoria, veja [1] e [2]. Com relação ao estudo de identificabilidade, estabelecem-se condições necessárias para assegurar tal característica. Em cada uma das situações (em termos de se conhecer ou não alguns dos parâmetros dos modelos) analisa-se a necessidade de se impor restrições em somente uma classe de parâmetros ou em ambas. No que diz respeito aos métodos de estimação, estes são feitos através do método de Máxima Verossimilhança, veja [1] e [2], considerando um determinado conjunto de parâmetros conhecido. Ou seja, primeiramente, considera-se o processo de estimação em que os parâmetros dos itens são conhecidos e depois os traços latentes conhecidos. Além da situação em que se desconhece todos os parâmetros. Nesse caso, explora-se o método de Máxima Verossimilhança Marginal (MVM) e o método por Máxima Verossimilhança Perfilada (MVP), para estimar os parâmetros dos itens, primeiramente, e em seguida os traços latentes, veja [5]. Em todos os casos estuda-se (via simulação) as principais propriedades dos métodos supracitados como: vício, variância e erro quadrático médio. A implementação computacional é feita no ambiente computacional, gráfico e estatístico **R**.

### 4 Resultados e discussões

Foi possível observar - além da complexidade natural do modelo - os problemas decorrentes à *falta de identificabilidade* fazem com que o processo de estimação seja uma tarefa delicada. Neste sentido, acredita-se que o estudo sobre as causas da *não-identificabilidade*, bem como as melhores estratégias para eliminá-la, seja uma importante contribuição deste trabalho. Outro fato de destaque foi a construção plena de um processo de estimação sob a perspectiva frequentista, com implementações computacionais escritas em linguagem **R**, viabilizando rotinas gratuitas, de código aberto e alta velocidade no processamento, essenciais à difusão de tais metodologias. Os resultados de simulação indicaram que todos os métodos recuperam bem os parâmetros, com uma ligeira vantagem ao se incorporar a restrição de *identificabilidade* ao final do processo iterativo. Além disso, depreendeu-se os fatores de maior impacto na precisão das estimativas dos traços latentes (NI e NC), dos parâmetros de discriminação (NE e NC) e parâmetros de dificuldade (NI, NC, NE e NI×NE).

## 5 Conclusões

No presente trabalho, num primeiro momento, a fim de melhor entender o significado dos parâmetros dos itens, avaliou-se graficamente como mudanças em seus valores influenciam o comportamento da CCI relacionada a cada categoria. Em relação à *identificabilidade*, condição necessária para garantir a estimabilidade do modelo, estudou-se, essencialmente de modo analítico, as condições para assegurar tal característica, com relação a aplicar restrições tanto nos parâmetros dos itens quanto nos traços latentes. Tais discussões serviram como base para a consecução do principal objetivo: desenvolver e implementar computacionalmente os métodos de estimação frequentista do MCPG. Através de estudos de simulação, avaliou-se a acurácia das estimativas considerando diversas situações de interesse prático. Com os resultados obtidos, esperamos satisfazer as necessidades de investigadores em vários campos de aplicação da TRI, disseminando uma maior utilização do MCPG. Em particular, na análise de dados de seleção de indivíduos e avaliação educacional.

## Referências

- [1] MURAKI, E. A generalized partial credit model: Application of an em algorithm. *Applied psychological measurement*, v. 16, n. 2, p. 159–176, 1992.
- [2] MURAKI, E. *A generalized partial credit model*. First. ed. New York, NY: Wim J. van der Linden and Ronald K. Hambleton eds, Springer-Verlag, 1997.
- [3] BAKER, F. B.; KIM, S.-H. *Item response theory : Parameter estimation techniques*. second. ed. New York, NY: MarcelDekker, Inc, 2004.
- [4] ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. *Teoria da resposta ao item : Conceitos e aplicações*. Primeira. ed. 14<sup>o</sup> SINAPE, ABE, 2000.
- [5] AZEVEDO, C. L. N. *Métodos de estimação na teoria da resposta ao item*. Dissertação de Mestrado, IME-USP, 2003.

# Utilização de Processos de Poisson Modulado por Markov na identificação de perfis de periculosidade na rodovia BR-381

Pedro Henrique Zampier<sup>1 3</sup>

Deive Ciro de Oliveira<sup>2 3</sup>

**Resumo:** O Brasil possui uma vasta extensão de malha rodoviária [4]. Isto se deve ao meio de transporte mais utilizado no país ser por veículos automotores [5]. Dentro deste contexto, verifica-se grande número de acidentes de trânsito, principalmente em rodovias federais. Estes provocam grande número de feridos e mortos, além de gerar despesas com tratamentos médicos, bem como com indenizações por danos materiais etc [4]. Políticas de manutenção das rodovias podem contribuir para a diminuição destes números. O presente trabalho busca, utilizando Processos de Poisson Modulados por Markov (MMPP), identificar, em uma rodovia, perfis de trechos com diferentes graus de periculosidade e estimar as taxas de acidentes em cada um destes trechos [9]. Foram utilizados dados públicos disponibilizados pelo DNIT (Departamento Nacional de Infraestrutura e Transporte) referente aos acidentes ocorridos entre os quilômetros 715 e 735 da rodovia federal BR-381, no período de 01 de janeiro a 31 de dezembro de 2011 [6].

**Palavras-chave:** *Processo de Poisson Modulado por Markov, estados de periculosidade, acidentes de trânsito, rodovias.*

## 1 Introdução

No Brasil, o principal meio de transporte utilizado é o terrestre feito por veículos automotores [5]. Devido ao país ter uma vasta extensão territorial, o principal meio de locomoção, tanto de pessoa quanto para o deslocamento de cargas e mercadorias, é feito através de rodovias. Segundo o boletim estatístico emitido pela CNT (Confederação Nacional de Transporte) em outubro de 2013, o Brasil possui um total de 1.584.712 km de malha rodoviária (pavimentada e não pavimentada), onde 77.981 km são rodovias federais. [4]. Nestas rodovias, o cenário acidentes é muito alto, e vem ocorrendo cada vez mais com maior frequência desde 2007, é o que demonstra o boletim informativo de acidentes da CNT, publicado em 2013 [4]. Identificar os trechos com alta taxa de acidentes é relevante pois, uma vez identificados, é possível verificar

---

<sup>1</sup>UNIFAL: Universidade Federal de Alfenas - Campus Varginha. Email: [zampier\\_ph@hotmail.com](mailto:zampier_ph@hotmail.com)

<sup>2</sup>UNIFAL: Universidade Federal de Alfenas - Campus Varginha. Email: [deive.oliveira@gmail.com](mailto:deive.oliveira@gmail.com)

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

o que causa estes acidentes. Fatores possíveis são o mal planejamento na construção das vias, a falta de sinalização adequada ou outros fatores.

## 2 Objetivo

O objetivo deste trabalho é analisar o conjunto de dados referentes a acidentes de trânsito ocorridos no trecho da rodovia federal BR-381 do quilômetro 715 ao quilômetro 735, buscando identificar perfis de periculosidade. Para isso foi utilizado o processo estocástico de Poisson Modulado por Markov.

## 3 Material e Metodologia

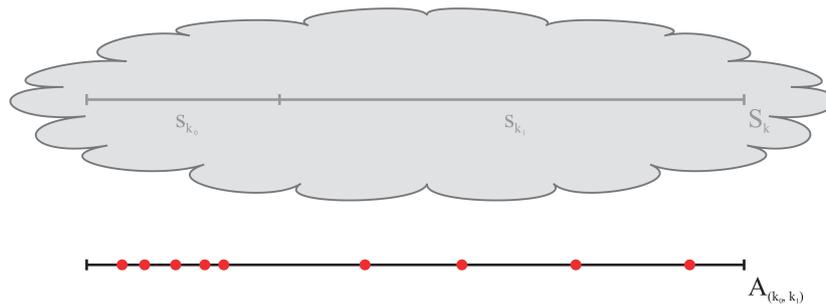
### 3.1 Processo de Poisson modulado por Markov

O modelo de Poisson modulado por Markov (*MMPP - Markovian Modulated Poisson Process*) é caracterizado por um processo estocástico observado  $A_{(k_0, k_1)}$  que segue uma distribuição de Poisson com parâmetros  $\lambda_{S_k}$ . Este é dependente de um processo não observável  $S_k$  que segue uma cadeia de Markov de espaço contínuo. [9] Tem-se que  $A_{(k_0, k_1)}$  refere-se ao número de acidentes no trecho  $[k_0, k_1]$  e  $S_k$  refere-se aos trechos de periculosidade da rodovia. O conjunto de parâmetros deste processo é dado por  $\theta = [\mathbf{Q}, \lambda_{S_k}]$ :

- $\mathbf{Q}$  é a matriz de transição de estados ou gerador infinitesimal. Nela os elementos da diagonal  $q_{jj}$  denotam a taxa de saída do estado  $S_j$ . Já os elementos fora da diagonal  $q_{ji}$  referem-se a taxa de transição do estado  $S_j$  para o estado  $S_i$
- $\lambda_{S_k}$  é o vetor de taxas de ocorrência de acidentes do processo observado dado por  $\lambda_{S_k} = [\lambda_1, \lambda_2, \dots, \lambda_k]$ , onde cada taxa de ocorrência  $\lambda_k$  do processo observado influencia na permanência em cada estado do processo não observado.

A Figura 1 ilustra um processo de Poisson modulado por Markov. Os pontos vermelhos representam as localidades onde ocorreram acidentes.

**Figura 1:** Processo de Poisson Modulado por Markov

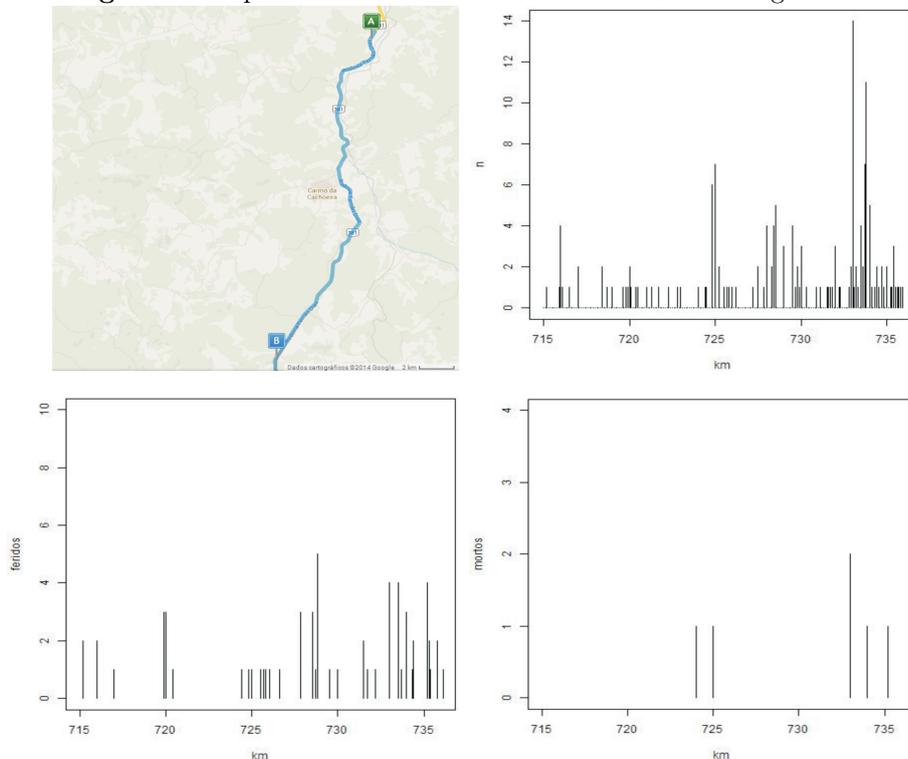


Utilizando o software R, por meio da biblioteca “HiddenMarkov”, foi obtida a estimação dos parâmetros via método da máxima verossimilhança. Foram ajustados 4 modelos considerando respectivamente de 2 a 5 perfis de trechos de periculosidade (estados).

### 3.2 Material

Foram utilizados dados de acidentes ocorridos nas rodovias federais no período de 01/01/2011 até 31/12/2011. [6]. A Figura 2 apresenta quatro imagens: A: Mapa do trecho estudado da Rodovia BR-381 do quilômetros 715 ao quilômetro 736; B: Histograma do número de acidentes por quilômetro; C: Histograma do número de feridos por quilômetro; D: Histograma do número de mortos por quilômetro.

**Figura 2:** Mapa da BR-381 do Km 715 ao Km 736 e Histogramas



Fonte: [www.mapa.guiamais.com.br](http://www.mapa.guiamais.com.br)

## 4 Resultado e Discussão

Através da análise descritiva, foram identificados os quilômetros 725 e 733 que apresentaram alto número de acidentes. No trecho total estudado ocorreram 166 acidentes, os quais ocasionaram 79 feridos e 6 óbitos durante todo o ano de 2011. Os parâmetros foram estimados via máxima verossimilhança utilizando-se o método numérico Baum-Welch [2]. Para escolha do melhor dentre os 5 modelos propostos, foi utilizado o critério de informação de Akaike [1] apresentado na tabela 1.

Tabela 1: AIC e verossimilhança absoluta.

Modelo	AIC	Verossimilhança
2 estados	466.2420*	229.1210*
3 estados	479.1955	230.5979
4 estados	509.8722	238.9361
5 estados	532.7486	241.3743

Fonte: Editado pelo autor.

O melhor modelo considerado foi o de 2 estados e seu vetor paramétrico estimado  $\hat{\theta} = [\hat{\mathbf{Q}}, \hat{\lambda}_{S_k}]$  foi:

$$\hat{\mathbf{Q}} = \begin{bmatrix} -0.39 & 0.39 \\ 4.37 & -4.37 \end{bmatrix}, \quad \hat{\lambda}_{S_k} = [4.71, 44.83].$$

## 5 Conclusão

Foram identificados 2 perfis de periculosidade no trecho analisado. No perfil mais seguro a taxa de acidentes foi de 4.71 por quilômetro. No mais perigoso, esta taxa cerca de 9 vezes maior (44.83). O tamanho médio dos trechos seguros é de 2.56 quilômetros e dos trechos perigosos de 0.23. Como futuros trabalhos, é possível inferir sobre a localização dos trechos de baixa e alta periculosidade utilizando o algoritmo de Viterbi [11].

## Referências

- [1] AKAIKE, Hirotugu. *Information theory and an extension of the maximum likelihood principle*. B. N. Petrov and F. Csaki (eds.), 2nd International Symposium on Information Theory: 267-81. Budapest: Akademiai Kiado. 1973.

---

\*Modelo considerado como o melhor modelo entre os ajustados, com AIC mais próximo de 0 e com menor verossimilhança absoluta.

- [2] BAUM, L.E.; PETRIE, T.; SOULES, G.; WEISS, N. *A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. The Annals of Mathematical Statistics*, 41(1), p. 164-171. 1970.
- [3] CASELLA, George; BERGER, Roger L. *Inferência estatística*. Trad. de Statistical inference, 2nd. ed. sob direção de Solange A. Visconte. São Paulo: Cengage Learning. 588 p. 2010.
- [4] CNT - CONFEDERAÇÃO NACIONAL DO TRANSPORTE. *Boletim Estatístico - CNT - Outubro 2013*. Disponível em: <<http://www.cnt.org.br>>. Acesso em: jan. 2014.
- [5] DENATRAN - Departamento Nacional de Trânsito, RENAVAM - Registro Nacional de Veículos Automotores. *Frota Nacional 2013*. Disponível em: <<http://www.denatran.gov.br>>. Acesso em: abr. 2014.
- [6] DNIT - Departamento Nacional de Infraestrutura e Transporte. *Estatística de Acidentes*. Disponível em: <<http://www.dnit.gov.br/rodovias/operacoes-rodoviaras/estatisticas-de-acidentes/>>. Acesso em: abr. 2014.
- [7] LIMA, Ieda Maria de Oliveira. et al. *Fatores condicionantes da gravidade dos acidentes de trânsito nas rodovias brasileiras*. Instituto de Pesquisa Econômica Aplicada (IPEA), 27 p. Brasília. 2008.
- [8] MULLER, Daniel. *Processos Estocásticos e Aplicações*. 22. ed. Coimbra: Almedina, 2007.
- [9] RYDÉN, Tobias. *An EM algorithm for estimation in Markov-modulated Poisson Processes*. Department of Mathematical Statistics, Lund University. Elsevier. Lund. Sweden. 1996. p. 431-447.
- [10] VALADAS, Rui. *Cadeias de Markov em tempo contínuo. Núcleo de Estudantes de Engenharia de Telecomunicação e Informática. 8 p. Lisboa. 2009. Disponível em: <<http://neerci.ist.utl.pt>>. Acesso em: dez. 2013.*
- [11] VITERBI, Andrew J. *Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. IEEE Transactions on Information Theory*. 1967. p. 260-269.

# *Choice-Based Conjoint Analysis: Um enfoque Bayesiano*

**Eduardo Campana Barbosa<sup>1</sup>**

**Carlos Henrique Osório Silva<sup>2</sup>**

**Moysés Nascimento<sup>3</sup>**

**Fabyano Fonseca e Silva<sup>4</sup>**

**Bruno Carlos Alves Pinheiro<sup>5</sup>**

**Resumo:** Neste trabalho demonstra-se o enfoque Bayesiano aplicado à metodologia *Choice-Based Conjoint Analysis*. Ilustra-se a modelagem, os resultados e inferências que podem ser obtidos por meio desta recente abordagem, sendo também apresentados os resultados do enfoque frequentista. Definiu-se como tratamento em estudo um tipo de camisa, composto por 4 fatores (Cor, Ilustração, Preço e Marca) e conclui-se que a maior probabilidade de escolha estava associada à camisa de cor preta, com ilustração, preço de R\$ 15,00 e com marca conhecida.

**Palavras-chave:** *Modelo Logit Multinomial, preferência do consumidor, escolha.*

## **1 Introdução**

A metodologia *Choice-Based Conjoint Analysis* (CBCA) é empregada em estudos da preferência do consumidor e utiliza a função Logit para modelar as probabilidades de escolha de diferentes tratamentos. Os tratamentos podem ser produtos, serviços ou conceitos e são definidos pela combinação de um nível de cada um dos atributos ou fatores considerados na análise. Tais estudos requerem um planejamento específico para a coleta de dados, que consistem na informação dos tratamentos escolhidos pelos consumidores participantes. Este trabalho conduzido com dados simulados apresenta um enfoque Bayesiano para CBCA.

## **2 Materiais e Métodos**

Definiu-se hipoteticamente que o tratamento em estudo seria uma camisa, descrita através de 4 fatores, cada um com 2 níveis, conforme apresentado na Tabela 1.

---

<sup>1</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: *duducampana@hotmail.com*

<sup>2</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: *chos@ufv.br*

<sup>3</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: *moysesnascim@ufv.br*

<sup>4</sup> DZO-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: *fabyanofonseca@ufv.br*

<sup>5</sup> DPP-UEMG: Universidade Estadual de Minas Gerais, Ubá-MG. Email: *bruno.pinheiro@uemgubá.edu.br*

Tabela 1 – Esquema com Fatores e Níveis do Produto Analisado (camisa)

Nível/Fator	1	2	3	4
	Cor	Ilustração	Preço	Marca
1	Preta	Sim	15,00	Conhecida
2	Branca	Não	20,00	Desconhecida

Ao combinar os níveis dos 4 fatores tem-se um total de  $J = 2^4 = 16$  tratamentos. Optou-se por um delineamento fatorial fracionado, com 8 tratamentos (Tabela 2) e apenas dos efeitos principais de cada fator. Para compor a variável resposta, definimos  $n_j$  onde  $j = 1, 2, 3, 4, 5, 6, 7$  e  $8$ , como o número de vezes que cada tratamento foi escolhido, conforme apresentado na Tabela 2.

Tabela 2 – Número de vezes que cada produto foi escolhido

Tratamentos (j)	2	3	5	8	10	11	13	16	Total
Número de Escolhas ( $n_j$ )	16	13	3	1	10	2	3	0	48

## 2.1 Modelo Logit Multinomial

Em um estudo composto por “ $r$ ” fatores, cada um com “ $m_r$ ” níveis, tem-se um conjunto de  $J = \prod_{s=1}^r m_s$  tratamentos distintos. A utilidade de um tratamento ou o benefício percebido por um consumidor, na *Choice-Based Conjoint Analysis* é descrito pelo modelo estatístico dado em (1).

$$U_{nj} = X_j \boldsymbol{\beta} + \varepsilon_{nj} \quad (1)$$

Em que  $U_{nj}$  é a utilidade atribuída pelo indivíduo  $n$  ao tratamento  $j$ ;  $X_j$  é a matriz com a codificação dos  $m_r$  níveis dos  $r$  fatores presentes no tratamento  $j$ ;  $\boldsymbol{\beta}$  é o vetor de parâmetros a serem estimados (um para cada fator em estudo);  $\varepsilon_{nj}$  é o erro aleatório associado à escolha do tratamento  $j$  pelo consumidor  $n$ , que pressupõe-se a seguir uma distribuição de valores extremos do tipo Gumbel [1].

Portanto, o consumidor irá escolher o tratamento  $j$ , se e somente se  $U_{nj} > U_{nk} \forall j \neq k; j, k \in \{1, 2, \dots, J\}$ , o que probabilisticamente e conforme detalhado em [1] resulta no modelo Logit Multinomial (2), onde  $P(Y_n = j | \mathbf{X})$  é a probabilidade de que o  $n$ -ésimo consumidor escolha o  $j$ -ésimo tratamento. A estimação de  $\boldsymbol{\beta}$  é pelo método da Máxima Verossimilhança Parcial, no pacote *survival* do *software* R, pela função *clogit*.

$$P(Y_n = j | \mathbf{X}) = \frac{e^{X_j \boldsymbol{\beta}}}{\sum_{k=1}^J e^{X_k \boldsymbol{\beta}}} \quad \forall j = 1, 2, \dots, J \quad (2)$$

## 2.2 Abordagem Bayesiana

Os métodos de inferência Bayesiana são governados pelo Teorema de Bayes, onde a distribuição a posteriori dos parâmetros,  $\pi(\boldsymbol{\beta})$ , é obtida combinando-se informações à priori com as informações provenientes da amostra aleatória, via função de verossimilhança,  $L(\boldsymbol{\beta})$ . Em geral:

$$\pi(\boldsymbol{\beta}) \propto L(\boldsymbol{\beta}) p(\boldsymbol{\beta})$$

No contexto da *Choice-Based Conjoint Analysis*, [2] sugerem o uso de uma distribuição normal  $r$ -variada  $\rightarrow \boldsymbol{\beta} \sim \text{Normal}_r(\bar{\boldsymbol{\beta}}, A)$ , com função densidade de probabilidade em (3):

$$p(\boldsymbol{\beta}) \propto |A|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})] \right\} \quad (3)$$

Em que  $\bar{\boldsymbol{\beta}}$  é o vetor que contém os valores médios da distribuição e  $A$  é o parâmetro que expressa sua precisão. Neste trabalho optou-se por uma priori pouco informativa para  $\boldsymbol{\beta}$ , conforme em (4), devido à ausência de informações na literatura sobre o tema.

$$\boldsymbol{\beta} \sim \text{Normal}_4 \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 0.01 \end{bmatrix} \right) \quad (4)$$

A verossimilhança é dada pelo produto das probabilidades conforme em (5). A combinação entre priori e verossimilhança resulta em uma forma complexa, logo, [2] sugerem uma aproximação normal assintótica para a densidade conjunta a posteriori (6), obtida pela expansão do log da verossimilhança em uma série de Taylor de segunda ordem, avaliada em  $\hat{\boldsymbol{\beta}}$  e  $H$  (Matriz Hessiana).

$$L(\boldsymbol{\beta}) = \prod_{n=1}^N \prod_{j=1}^J \left( \frac{e^{x_j \boldsymbol{\beta}}}{\sum_{k=1}^J e^{x_k \boldsymbol{\beta}}} \right)^{n_j} \quad (5)$$

$$\pi(\boldsymbol{\beta}) \propto |H|^{1/2} \exp \left\{ \frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T H (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\} \quad (6)$$

Para gerar amostras das distribuições a posteriori marginais dos parâmetros empregou-se o algoritmo MCMC *Independence Metropolis Hasting*, que tem como distribuição candidata uma  $t$  de Student  $r$ -variada, com  $\nu$  graus de liberdade, parâmetros  $\mu = \hat{\boldsymbol{\beta}}$  e  $A = H^{-1}$  e função densidade (7).

$$p(\boldsymbol{\beta} | \nu, \hat{\boldsymbol{\beta}}, H^{-1}) \propto |H|^{\frac{1}{2}} \left[ \nu + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T H (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right]^{-\frac{\nu+r}{2}} \quad (7)$$

Destaca-se que a aproximação mencionada em (6) esta contida implicitamente em (7), porém, com algumas modificações e a inclusão do parâmetro  $\nu$  para tornar a distribuição candidata uma  $t$  de Student, mais densa e com caudas mais gordas do que a distribuição de interesse. Este procedimento esta disponibilizado no pacote *bayesm* do *software* R, via função *rmnlIndepMetrop*.

### 3 Resultados e Discussão

A Tabela 3 apresenta as estimativas frequentistas ( $\hat{\boldsymbol{\beta}}$ ) e Bayesianas (média a posteriori) para o vetor de parâmetros do modelo Logit Multinomial. Para a segunda utilizou-se 70.000 iterações, com um *burn-in* de 4.000 e *thin* de 5.

Tabela 3 – Resultados Frequentistas [Erro Padrão] e Bayesianos [Desvio-Padrão a Posteriori]

Método	Cor	Ilustração	Preço	Marca
Frequentista	0,6209 [0,5639]	1,1709* [0,5639]	2,0872* [0,5639]	0,7885* [0,3114]
Bayesiano	0,8366 [0,6798]	1,3964 [0,6798]	2,3675 [0,6798]	0,8081 [0,3224]

\* Significativo a 5% de probabilidade

As estimativas obtidas pelos enfoques Bayesiano e Frequentista foram muito parecidas. O fator cor não foi significativo a 5% de probabilidade e, portanto, não influenciou na escolha do consumidor. Após a estimação de  $\beta$ , é possível calcular a probabilidade de escolha de todos os tratamentos, inclusive dos não avaliados pelos participantes do estudo (retirados no delineamento fatorial fracionado), sendo este o diferencial da metodologia. A Tabela 4 apresenta os seis tratamentos que obtiveram as maiores probabilidades de escolha.

Tabela 4 – Probabilidades de Escolha dos tratamentos

Tratamento	Cor	Ilustração	Preço	Marca	Probabilidades	
					Frequentista	Bayesiano
1	Preto	Sim	15	Conhecida	0,3037	0,3537
2	Branco	Sim	15	Conhecida	0,1632	0,1532
9	Preto	Sim	15	Desconhecida	0,1380	0,1577
3	Preto	Não	15	Conhecida	0,0942	0,0875
10	Branco	Sim	15	Desconhecida	0,0742	0,0683
4	Branco	Não	15	Conhecida	0,0506	0,0379

As probabilidades estimadas pelo método bayesiano foram também muito próximas das frequentistas. Pela Figura 1 parece ter ocorrido a convergência das cadeias, o que é comprovado pelo critério de Geweke a 5% de probabilidade ( $p$ -valor > 0,05), conforme na Tabela 5. Na Figura 2 encontra-se um esboço das densidades a posteriori estimadas. Tais resultados foram obtidos pelo pacote *BOA* do *software* R.

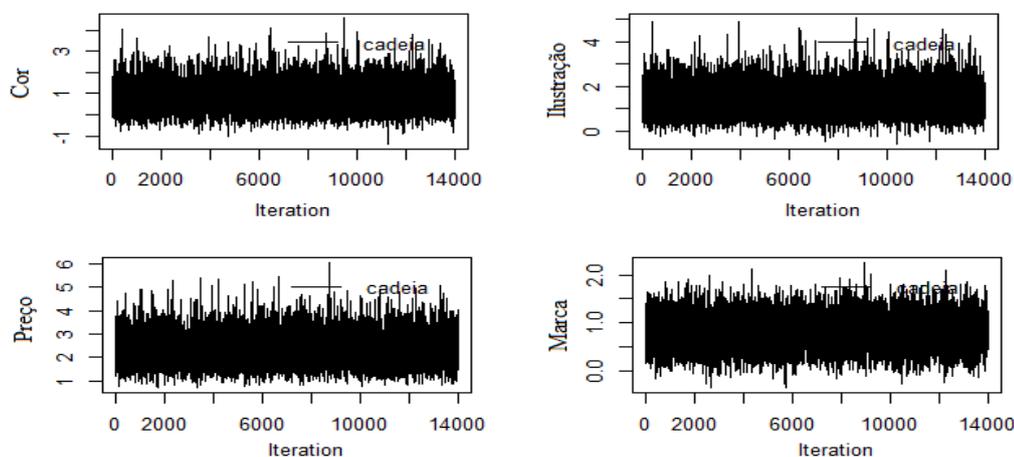


Figura 1 – Ilustração das Cadeias de Markov.

Tabela 5 – Resultados para o Critério de Geweke

Estadísticas	Cor	Ilustração	Preço	Marca
Z-Score	0,1402	0,4035	0,0615	-0,9265
Valor-p	0,8884	0,6865	0,9509	0,3541

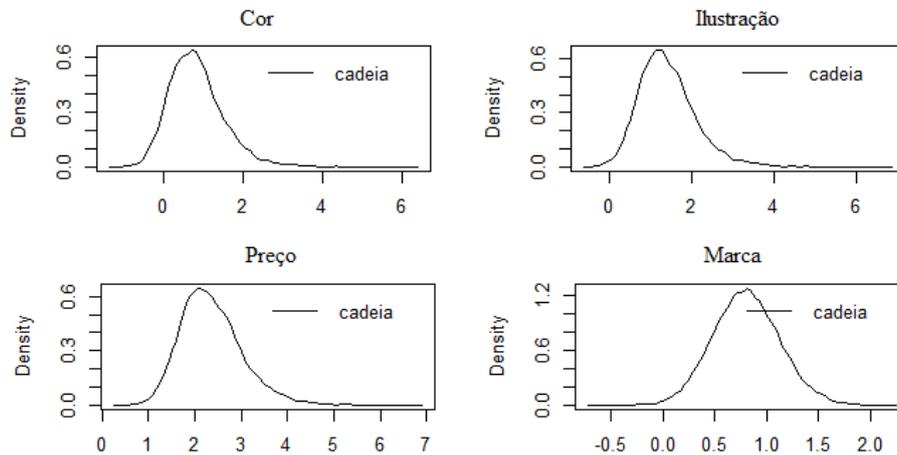


Figura 2 – Esboço das densidades a Posteriori marginais estimadas

## 4 Conclusões

O enfoque Bayesiano mostrou-se adequado para aplicação em estudos de CBCA, pois as estimativas obtidas por meio da média a posteriori, bem como as probabilidades de escolha de cada tratamento, foram bem próximas das frequentistas. O tratamento com maior probabilidade de escolha foi o de cor preta, com ilustração, preço R\$ 15,00 e com marca conhecida. Sugere-se que novos trabalhos e pesquisas com o enfoque Bayesiano sejam desenvolvidos, no intuito de obter mais informações, principalmente sobre a questão de distribuições a priori informativas. Além disso, existe a opção de se implementar tal metodologia de uma forma mais flexível utilizando-se o *software Winbugs*, o que poderia oferecer resultados interessantes.

## 5 Referências Bibliográficas

- [1] McFadden, D. (1974), “**Conditional Logit Analysis of Qualitative Choice Behavior**”, *Frontiers in Econometrics*, Academic Press, P. Zarembka Eds., New York, pp. 105-142.
- [2] Rossi, E., Allenby, G., McCulloch, R. **Bayesian statistics and marketing**, John Wiley & Sons, 2005.

# Retorno de ações: uma análise econométrica de companhias de capital aberto no período de 2008 a 2010

Maria Eugenia de Oliveira Batista<sup>1 4</sup>

Gabriel Rodrigo Gomes Pessanha<sup>2 4</sup>

Jennifer Elis Pires<sup>3 4</sup>

**Resumo:** O objetivo desta pesquisa é desenvolver um modelo alternativo para o Capital Asset Pricing Model (CAPM), de modo a analisar o efeito do risco de investimento, do tamanho da empresa, da atividade comercial e de mercado. Esta análise centra-se na relação entre o retorno, *Value at Risk* (VaR) e risco de mercado; e na relação entre o retorno, tamanho, liquidez e endividamento. Foi empregado um modelo de regressão com dados em painel para análise de dados. As amostras são compostas por 5 ações de alta liquidez da BM&Fbovespa para o período de 2008-2010. Os resultados demonstraram que o modelo construído a partir do logaritmo do retorno das ações depende em 97,66% dos casos das variáveis em estudo, com um nível de significância para teste inferior a 1%, sendo desta maneira um estudo altamente confiável. Esse estudo é inovador, visto que procura encontrar soluções alternativas para complementar o modelo CAPM, ou seja, procura um melhor modelo que se adapte a realidade do mercado financeiro atual. Diante disto, esse trabalho possui grande importância no cenário econômico devido a grande relevância que a gestão de risco e o modelo CAPM têm nos dias de hoje.

**Palavras-chave:** CAPM, regressão, retorno.

## 1. Introdução

A mensuração do risco, inerente aos mercados financeiros e comumente associado ao desvio padrão passado dos retornos do ativo, levou ao desenvolvimento de um dos mais clássicos modelos usados em Finanças: o capital asset pricing model, ou modelo de precificação de ativos financeiros, ou, simplesmente CAPM. O modelo CAPM supõe a existência de mercados eficiente, onde os preços dos ativos refletem o consenso geral sobre todas as informações disponíveis. Tosta de Sá (1999) interpreta que o beta nada mais é do que o coeficiente de regressão da reta que melhor ajusta o retorno de um título com o ganho da carteira do mercado; ou seja, é um indicador que mede como reage o preço de um título às oscilações do índice representativo de seu mercado.

De acordo com Fama e French (1992) a teoria CAPM só leva em consideração a relação entre o rendimento esperado com risco de mercado (beta), mas existem vários outros fatores importantes que também estão envolvidos com o rendimento esperado do mercado financeiro. Diante disso, o

---

<sup>1</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha-MG. Email: [mag\\_batista@hotmail.com](mailto:mag_batista@hotmail.com)

<sup>2</sup> ICSA – UNIFAL: Universidade Federal de Alfenas Varginha-MG. Email: [gabriel.pessanha@unifal-mg.edu.br](mailto:gabriel.pessanha@unifal-mg.edu.br)

<sup>3</sup> ICSA – UNIFAL: Universidade Federal de Alfenas-MG. Email: [jenni.elis@hotmail.com](mailto:jenni.elis@hotmail.com)

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

problema deste estudo é associar uma solução alternativa ao método de precificação de ativos que existe hoje.

Entretanto, recentes testes têm evidenciado a existência de outros fatores explicativos dos retornos esperado no mercado de ações, que não estão presentes no CAPM, os teóricos de finanças ainda parecem longe de obter consenso acerca da utilidade do CAPM como modelo explicativo da formação dos preços dos ativos. Assim, cria-se a necessidade de estudar outros fatores que possam afetar os retornos de investimentos em ações, de modo a tentar encontrar um modelo alternativo e complementar ao modelo CAPM.

## 2. Objetivo

Este estudo tem como objetivo geral encontrar soluções alternativas para complementar o modelo CAPM, de modo que irá utilizar o Value at Risk (VaR) como uma das variáveis para estimar a correlação entre o risco e o retorno. Assim, medirá o quão significativo é a relação entre Value at Risk (VaR) com os retornos das ações como um método alternativo de precificação de ativos. Este estudo também determinará se, o endividamento, a liquidez e o tamanho das empresas, podem ser usados como uma medida alternativa de investimento de risco ao lado do risco de mercado (beta) para determinar a precificação de ativos no mercado de capitais.

## 3. Metodologia

A fim de se testar um modelo alternativo que analisa o efeito do risco de investimento, da atividade comercial e de mercado este estudo utiliza dados trimestrais de ações que estão listadas na BM&FBOVESPA, a partir do primeiro trimestre do ano de 2008 até o terceiro trimestre de 2010. As amostras de pesquisa são de 5 ações de alto grau de negociação ativas negociadas na BM&FBOVESPA: BBDC4 (Bradesco), GGBR4 (Gerdau), ITSA4 (Itausa), ITUB4 (Itaú Unibanco) e PETR4 (Petrobras). A estrutura conceitual/teórica é apresentada na Figura 1.

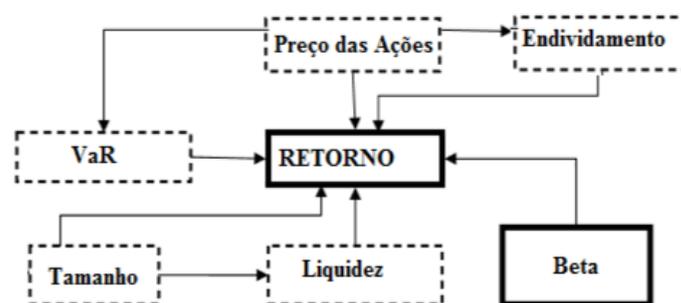


Figura 1: Quadro do Modelo Conceitual Teórico

Para analisar a influência dos eventos VaR, beta, tamanho, liquidez e endividamento sobre o retorno das ações estudada, foi utilizada a metodologia de regressão. O modelo de regressão linear, segundo Krajewski, Ritzman e Malhotra (2009), é um dos modelos causais mais conhecidos e utilizados, que consiste de uma variável chamada de dependente estar relacionada a uma ou mais

variáveis independentes por uma equação linear. A equação de regressão do modelo ajustado é apresentada a seguir:

$$R_{i,t} = \delta_t + \gamma_{1,t}VaR(\alpha)_{i,t} + \gamma_{2,t}\beta_{i,t} + \gamma_{3,t}Tam_{i,t} + \gamma_{4,t}Liq_{i,t} + \gamma_{5,t}End_{i,t} + \gamma_{6,t}Re_{i,t} + \varepsilon_i \quad (1)$$

em que:

$R_{i,t}$ : O nível do log do retorno das ações  $i$  no período  $t$ .

$VaR(\alpha)_{i,t}$ : A perda máxima em determinado período de tempo (horizonte alvo) com um certo nível de confiança das ações  $i$  no período  $t$ .

$\beta_{i,t}$ : Risco de Mercado das ações  $i$  no período  $t$ .

$Tam_{i,t}$  Tamanho ou volume das ações  $i$  negociadas no período  $t$ .

$Liq_{i,t}$ : Liquidez das ações  $i$  no período  $t$ .

$End_{i,t}$ : Endividamento das ações  $i$  no período  $t$ .

$Re_{i,t}$ : O nível do retorno das ações  $i$  no período  $t$ .

#### 4. Análise e discussão

Partindo para a análise econométrica dos dados, a correlação foi elaborada com o objetivo de se verificar a existência de multicolinearidade que pudesse comprometer o ajuste final do modelo de regressão. A partir dos resultados encontrados, pode-se afirmar que não existe comportamento de multicolinearidade entre as variáveis estudadas.

Para testar a veracidade do novo modelo, e a relação das variáveis com o log do retorno foi feita uma regressão, alcançando deste jeito um modelo alternativo ao CAPM significativo. De forma geral, pode-se afirmar que o modelo econométrico apresentou um ótimo ajustamento, uma vez que os testes demonstraram alta significância dos parâmetros estimados (Tabela 1).

Tabela 1: Resultados da regressão linear múltipla

	Coefficiente	Erro padrão	Teste t
Constante	0,120991	0,0555161	0,03483**
LC	-0,0346794	0,00898138	0,00037***
LS	0,448971	0,0175201	0,00804***
Retorno	-0,0194955	0,0104528	<0,00001***
Tamanho	0,0589772	0,00684438	0,00671***
Ind_End_Geral	0,0589772	0,0231338	0,01444**
l_Beta	-0,00250849	0,00146626	0,09432*
l_VaR	-0,00222419	0,00374621	0,55581
Coeficiente de determinação ( $R^2$ ) 0,976698***			

Legenda: \* significativo a 10%; \*\* significativo a 5% ; \*\*\* significativo a 1%.

Esse modelo estudou a relação da variável dependente log dos retornos e as variáveis independentes: liquidez corrente (LC), liquidez (LS), retorno, tamanho, índice de endividamento geral (Ind\_End\_Geral), log do beta (l\_Beta) e o log do VaR ( l\_VaR). A regressão mostra que o log do retorno das ações depende em 97,66% dos casos das variáveis em estudo, com um nível de significância, teste F, inferior a 1%, sendo desta maneira um ajuste altamente confiável.

Partindo para a análise individual das variáveis, pode-se afirmar que existe uma relação inversa com a variável Liquidez Corrente (LC) das empresas. Além disso, a regressão revela que se a liquidez corrente tem um aumento na ordem de 100%, o log do retorno diminuirá 3,46%, tal estatística é confirmada pelo teste t com um nível de significância inferior a 1%. Esse resultado não era esperado, já que a expectativa era de que um aumento na liquidez influenciaria positivamente os retornos auferidos.

Comportamento inverso pode ser observado com a variável liquidez seca (LS), mostrando que um aumento de 100% da liquidez seca fará o log do retorno subir 4,86%, com um nível de significância inferior a 1%. Esse dado é significativo, pois se espera que uma empresa que está em bom estado de liquidez tenha maiores retornos. Assim, os indicadores de liquidez, que avaliam a capacidade de pagamento da empresa em relação às suas obrigações, tiveram resultados diferentes.

O retorno tem relação positiva com o log do retorno, sendo que se o retorno sobe 100% o log do retorno também aumentará em 44,89%. Esses dados são significativos a 1%, o que os confirma. Esse resultado é mais que esperado, tendo em vista que só mudou a metodologia de cálculo do retorno então a dependência de um para o outro deve ser positiva e com valor alto.

Quanto ao tamanho da empresa, ela possui uma relação negativa sobre o log do retorno, se o tamanho da empresa crescer 100% o log do retorno diminuirá numa razão de 1,94 %, no qual esses valores são significativos a 1%. Esse resultado refuta a hipótese de que o tamanho da empresa influencia de maneira positiva o retorno das ações.

Observa-se que a variável índice de endividamento geral (Ind\_End\_Geral) tem um efeito positivo sobre o retorno de fechamento das ações, já que se o índice aumentar 100% o log do retorno aumentará 5,89%. A veracidade dos dados se comprova com significância a 5%, tal constatação contraria as expectativas iniciais.

O log do beta (l\_Beta) apresentou uma relação negativa com o log do retorno, uma vez que um aumento na ordem de 100% do Beta fará o log do retorno reduzir em 0,2%, tal coeficiente pode ser ratificado pelos testes t com significância inferior a 10%. Esse resultado não era esperado, tendo em vista que um dos pressupostos básicos das finanças corporativas se refere a relação direta entre risco e retorno. Já o Log do VaR(l\_VaR) não apresentou significância estatística no modelo, o que não era esperado, pois o VaR é uma medida de risco e deveria apresentar alguma influência sobre o retorno das ações.

## 5. Considerações finais

Hoje em dia a gestão de risco de mercado vem se tornando cada vez mais difícil, isso ocorre pelo fato de que o mercado está cada vez mais complexo, no qual ele passa a ser influenciado por diversos fatores diferentes. O retorno no mercado financeiro não necessariamente engloba apenas o risco envolvido nas operações. Diante disto, neste trabalho foi apontado outras variáveis que afetam significativamente o retorno.

O modelo alternativo ao modelo CAPM criado neste trabalho não conseguiu atender o objetivo principal do trabalho, usar o VaR como metodologia de risco alternativa, contudo foi criado um novo modelo consistente que envolve outras variáveis financeiras além do risco.

Desta forma, chegou-se ao resultado que as variáveis independentes: liquidez corrente, liquidez seca, tamanho, retorno, log do beta, índice de endividamento e retorno promove efeitos significativos no retorno das ações pesquisadas. De maneira que a liquidez corrente, o tamanho da empresa e o log do beta possuem uma relação negativa do retorno, enquanto as demais uma correlação positiva.

Pela grande relevância que a gestão de risco e o modelo CAPM tem nos dias de hoje, sugere-se analisar um período maior de tempo, de modo a possuir mais dados para criar um modelo ainda mais significativo. Assim como, tentar ajustar o modelo para que se consiga utilizar a medida de risco VaR como medida alternativa ao beta e ver qual melhor modelo que se adapta a realidade do mercado financeiro da atualidade.

## Referências

- [1] FAMA, E.F.; FRENCH, K.. The Cross-Section of Expected Stock Returns, *Journal of Finance*, 47 (2), (June, 1992), 427-465.
  
- [2] KRAJEWSKI, L. J.; RITZMAN, L. P.; MALHOTRA, M. Administração da produção e operações. 8. ed. São Paulo: Pearson Prentice Hall, 2009.
  
- [2] TOSTA DE SÁ, Geraldo. *Administração de investimentos, teoria de carteiras e gerenciamento de risco*. Rio de Janeiro: Qualitymark, 1999.

# Previsão das Operações de Fusões e Aquisições: Redes Neurais Artificiais Aplicadas ao Setor Financeiro Brasileiro

Jennifer Elis Pires<sup>1 3</sup>

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

**Resumo:** O objetivo deste trabalho é desenvolver um modelo de previsão das operações de fusões e aquisições (F&A) para instituições financeiras brasileiras utilizando Redes Neurais Artificiais (RNA). Diversos estudos apontam que as operações de F&A influenciam no valor das empresas, e muitos destes indicam que a grande parte dos ganhos nessas operações se volta para os acionistas das empresas alvo. A possibilidade de identificar a ocorrência dessas operações é atrativa a empresas e investidores, pois estes poderão usar de informações diferenciadas na tomada de decisão e optar pelo investimento de melhor retorno. A perspectiva de ganhos anormais, oriundos de uma informação relevante, é um assunto altamente discutido na teoria de finanças. Utilizando a técnica de RNA, o presente estudo trabalha com 16 indicadores, divididos em seis dimensões (adequação de capital, qualidade dos ativos, eficiência, rentabilidade, liquidez e características gerais), para construir uma RNA. A partir dos resultados foi possível observar que a RNA construída classificou corretamente, aproximadamente, 89% das instituições que passaram ou não por operações de F&A. Desta forma, espera-se com este trabalho contribuir para a escassa literatura e para as discussões acerca da importância de modelos de previsão das operações de F&A no setor bancário e em mercados emergentes.

**Palavras-chave:** *Fusões e aquisições, instituições financeiras, redes neurais.*

## 1. INTRODUÇÃO

A economia global passou e passa por diversas transformações que intensificam a cada dia a economia. Todos os atores econômicos estão sujeitos às consequências dessa transformação, e, esta reflete diretamente no ambiente de negócios, seja em momentos de ascensão ou de recessão econômica. O Brasil, enquanto mercado emergente tem passado por diversas transformações visando conquistar crescimento sólido e lugar de destaque na economia mundial. O país atualmente é considerado um mercado atraente para os investidores estrangeiros devido a uma variedade de fatores, sobretudo, por ter conquistado relativa estabilidade econômica e política, ter controlado a inflação e consistir em um mercado consumidor em expansão.

Todas essas transformações na economia nacional auxiliaram a ampliação do número de transações de fusões e aquisições (F&A). As empresas nacionais e estrangeiras vislumbram nessas

---

<sup>1</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha-MG. Email: [jenni.elis@hotmail.com](mailto:jenni.elis@hotmail.com)

<sup>2</sup> ICSA – UNIFAL: Universidade Federal de Alfenas Varginha-MG. Email: [gabriel.pessanha@unifal-mg.edu.br](mailto:gabriel.pessanha@unifal-mg.edu.br)

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

operações oportunidades de crescimento, ademais, tais operações podem ser consideradas resultado das estratégias da empresa (ÖBERG; HOLTSTRÖM, 2006).

Considerada uma forma de atividade que pode proporcionar aos participantes do mercado o controle corporativo, ganho de poder de mercado, dentre outros, as operações de F&A ainda são um tema de constantes discussão e divergências entre profissionais, governos e sociedade. Parte das discussões, principalmente no mercado e entre os teóricos, é sobre a possibilidade de previsão destes acontecimentos.

Considerando o volume de transações de F&A envolvendo empresas brasileiras no setor bancário e importância de tal estratégia de crescimentos nos resultados das companhias, torna-se relevante estudar e analisar a existência de padrões de ocorrência destes processos e desenvolver metodologias de previsão a partir de variáveis que expressam as características gerais, econômicas e financeiras da firma e que são capazes de influenciar a ocorrência do processo de fusão e/ou aquisição. Sendo assim, o objetivo principal deste trabalho é desenvolver um modelo de previsão de operações de fusões e aquisições para instituições financeiras brasileiras utilizando a metodologia de Redes Neurais Artificiais.

## 2. METODOLOGIA

A amostra é composta de empresas que foram e não foram alvos de fusão e aquisição no período de 1999 a 2013, os dados são trimestrais e totalizam 900 observações (60 observações para cada instituição financeira estudada). Os dados utilizados nas análises estatísticas foram obtidos através do banco de dados Economatica. O *software* utilizado para o tratamento dos dados foi o *Neuro Solutions* versão 4.20. As instituições financeiras selecionadas para esta pesquisa são: Banco do Brasil, ItauUnibanco, Bradesco, Banrisul, Nordeste do Brasil, Porto Seguro, Banestes, Mercantil do Brasil, Alfa, Banco BRB, Amazonia, Banpara, Banese, Finansinos, Companhia de Seguros Aliança da Bahia.

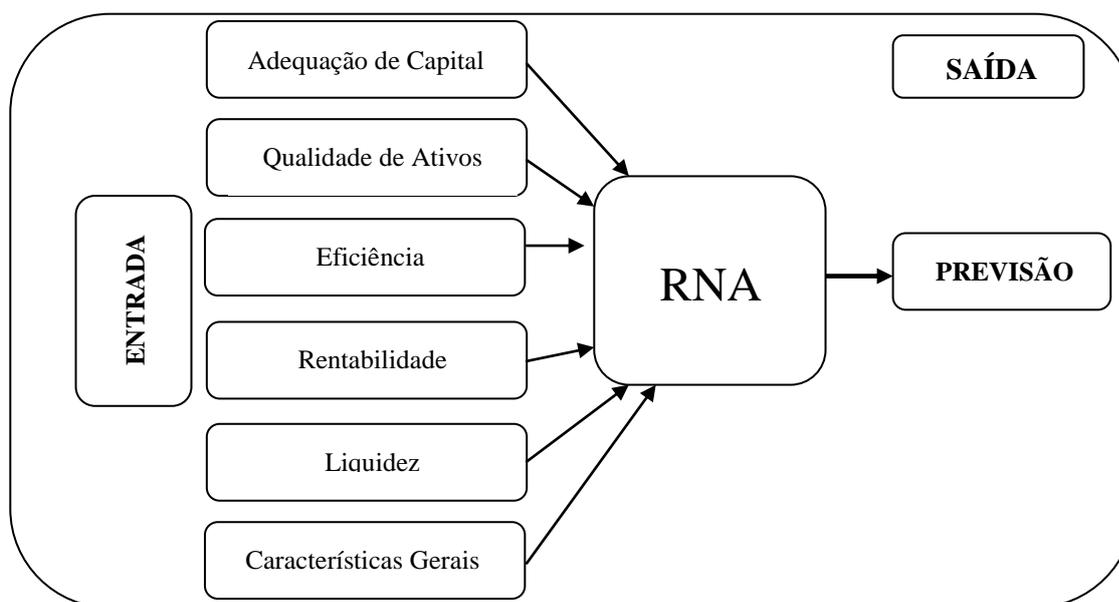


Figura 1: Modelo conceitual-teórico - Diagrama da rede proposta

Os indicadores utilizados nesta pesquisa seguem a estrutura do Plano de Contas – COSIF. Foram realizadas adaptações a estes indicadores a fim de adequá-los ao sistema bancário e em função também do nível de contas contábeis divulgadas pelo BACEN. Assim, utilizou-se 16 indicadores os quais foram divididos em seis dimensões, conforme apresentado na Figura 1.

### 3. ANÁLISE E DISCUSSÃO

Dentre os diversos tipos de RNA, optou-se pela aplicação da retro-propagação por ser este o mais utilizado para a previsão desta natureza. As redes construídas durante o tratamento dos dados são constituídas por uma camada de entrada, uma camada intermediária e uma camada de saída (previsão).

Segundo Vieira e Cavalcanti (2010), o processo de construção de um modelo de rede neural pode ser dividido em duas etapas, a primeira fase é conhecida como fase de aprendizagem ou de treinamento, nesta fase as conexões entre os neurônios são ajustadas até que o modelo seja capaz de compreender as relações entre a (s) variável (is) independente (s) e a (s) variável (is) dependente (s) com o grau de precisão desejado. A segunda fase, fase de teste ou validação, avalia-se a qualidade do modelo ajustado por meio da precisão das previsões *ex post*.

Seguindo as diretrizes apresentadas no trabalho de Castro Júnior (2003) e considerando as orientações de Gately (1996), o qual sugere a utilização da maioria dos dados disponíveis para o processo de aprendizagem e o restante na fase de validação, a amostra de 900 observações foi dividida em três subgrupos. O primeiro subgrupo é a amostra de treinamento que foi composta por 90 observações (60 observações de instituições alvos de F&A e 30 observações de instituições não alvos de F&A). O segundo subgrupo é a amostra de validação e é composta por 630 observações (480 observações de instituições alvos de F&A e 150 observações de instituições não alvos de F&A) e, por último, o terceiro subgrupo, a amostra de teste, foi composto de 180 observações, das quais 120 são de instituições alvos de F&A e 60 de instituições não alvos de F&A (Tabela 1).

Tabela 1: Subgrupos utilizados na construção da RNA

Subgrupos	Observações		Porcentagem		Total de Observações	Porcentagem em relação ao total
	Alvos de F&A	Não Alvos de F&A	Alvos de F&A	Não Alvos de F&A		
<b>Treinamento</b>	60	30	66,67%	33,33%	90	10%
<b>Validação</b>	480	150	76,20%	23,80%	630	70%
<b>Teste</b>	120	60	66,67%	33,33%	180	20%
<b>TOTAL</b>	<b>660</b>	<b>240</b>	<b>73,33%</b>	<b>26,67%</b>	<b>900</b>	<b>100%</b>

Na Tabela 2 apresenta-se um dos indicadores mais importantes da qualidade de ajuste e aprendizagem da rede neural artificial. Na fase de treinamento, a análise do MSE revela a qualidade do aprendizado e adaptação da rede para a previsão da variável dependente (alvo ou não alvo de fusões e aquisições). Para a amostra de treinamento, o valor mínimo obtido para a média dos MSE's foi menor que 0,0001, já para a amostra de validação o valor encontrado foi de, aproximadamente, 0,022, este valor pode ser considerado aceitável, já que a variável de saída (dependente) é uma variável binária

que assume valor 0 e 1 e, por isso, valores muito alto para o MSE poderia representar indícios de problemas no ajuste da RNA.

Tabela 2: Valores mínimos e desvio padrão para o MSE para as amostras de treinamento e validação

	<b>Mínimo (Treinamento)</b>	<b>Desvio Padrão (Treinamento)</b>	<b>Mínimo (Validação)</b>	<b>Desvio Padrão (Validação)</b>
<b>Média dos MSE's Mínimos</b>	0,000097845	0,00018971	0,02189971556	0,025900167
<b>Média dos MSE's Finais</b>	0,000097845	0,00018971	0,02769910121	0,0147716011

Na Tabela 3 apresentam-se os resultados de classificação das instituições financeiras estudadas.

A rede neural artificial ajustada foi capaz de classificar corretamente, na fase de treinamento, cerca de 75,83% de todas as instituições financeiras incluídas na amostra. Na amostra de validação a RNA classificou corretamente, aproximadamente, 85,54% das observações. Para a amostra de teste, que mede a real capacidade de generalizações da rede desenvolvida, os dados são promissores. A rede classificou corretamente, entre instituições alvos e não alvos de fusões e aquisições, cerca de 89% dos dados analisados.

A rede foi construída com todas as variáveis selecionadas para o estudo (16 variáveis), todavia, uma análise de sensibilidade para a seleção das variáveis e para a identificação da quantidade adequada de neurônios na camada intermediária poderia eliminar e/ou amenizar a multicolinearidade e melhorar os resultados da classificação da rede neural.

Tabela 3: Resultados da classificação obtida pela RNA

	<b>Treinamento</b>		<b>Validação</b>		<b>Teste</b>	
	<b>Alvos de F&amp;A</b>	<b>Não-Alvos de F&amp;A</b>	<b>Alvos de F&amp;A</b>	<b>Não-Alvos de F&amp;A</b>	<b>Alvos de F&amp;A</b>	<b>Não-Alvos de F&amp;A</b>
<b>Total</b>	60	30	480	150	120	60
<b>Classificação Correta</b>	49	21	402	131	103	55
<b>% de Acertos</b>	81,67%	70,00%	83,75%	87,33%	85,83%	91,67%

De maneira geral, pode-se afirmar que os resultados encontrados durante o desenvolvimento deste trabalho estão coerentes com as evidências encontradas por Adelaja, Nayga e Farooq (1999), os quais concluíram que a liquidez da empresa, rentabilidade, alavancagem e os dividendos gerados são características que podem determinar um alvo de F&A. Os resultados aqui encontrados também estão em sintonia com os resultados de Changqi e Ningling (2010), os quais afirmaram que o desempenho das empresas antes das F&A é essencial para o acontecimento das mesmas.

## 5. CONSIDERAÇÕES FINAIS

Tendo em vista a escassez de estudos empíricos sobre fusões e aquisições em instituições financeiras e a ausência da aplicação de modelos estatísticos para a previsão da ocorrência de fusões e aquisições, o objetivo deste trabalho foi o de desenvolver um modelo de previsão através das Redes Neurais Artificiais para as operações de F&A em instituições financeiras no mercado brasileiro.

As variáveis de entrada da rede foram definidas com base em 16 indicadores, os quais foram divididos em seis dimensões. Os indicadores seguem a estrutura do Plano de Contas – COSIF e foram necessárias adaptações a estes indicadores a fim de adequá-los ao sistema bancário e em função também do nível de contas contábeis divulgadas pelo BACEN e pelo Economática.

A partir destes 16 indicadores divididos em seis dimensões (adequação de capital, qualidade dos ativos, eficiência, rentabilidade, liquidez e características gerais), construiu-se uma rede neural artificial para a classificação das instituições financeiras estudadas. A RNA ajustada foi capaz de classificar corretamente, aproximadamente, 86% das instituições financeiras que passaram por fusões e aquisições e, aproximadamente, 92% das instituições financeiras que não optaram por estratégias de F&A. De maneira geral, o modelo apresentou na fase de teste (etapa de generalizações) um grau de acerto de, aproximadamente, 89%. Os dados são promissores, todavia as limitações estatísticas da metodologia precisam ser consideradas.

Por fim, ressalta-se a importância do desenvolvimento de trabalhos como este para mercados emergentes. Modelos de previsão podem trazer mais segurança e amenizar o risco assumido pelos investidores. Além disso, podem fornecer úteis informações para a tomada de decisão empresarial, uma vez que elenca variáveis importantes para a classificação de empresas alvo e não alvo de F&A. Espera-se, com este trabalho, contribuir para as discussões acerca da importância de modelos de previsão de F&A no setor bancário e em mercados emergentes. Para pesquisas futuras sugere-se estudo de sensibilidade nas variáveis de entrada incluídas no modelo. Além disso, tendo em vista o fato de que qualquer estatística de previsão é passível de erros, sugere-se também a utilização de outras metodologias e técnicas para a complementação dos resultados. Acredita-se que a aplicação destas sugestões pode aumentar ainda mais a robustez e o poder de classificação das instituições que passaram ou não por operações de fusões e aquisições.

## Referências

- [1]ADELAJA, A., NYAGA, R., FAROOQ, Z., Predicting Mergers and Acquisitions in the Food Industry. *AGRIBUSINESS* 15, (1), 1-23, 1999.
- [2]CASTRO JUNIOR, F. H. F. de. *Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais*. São Paulo: FEA/USP, 2003. 169 p.
- [3]CHANGQI, W.; NINGLING, X. *Determinants of Cross-Border Merger & Acquisition Performance of Chinese Enterprises*, *Procedia - Social and Behavioral Sciences*, 2(5), 6896-6905, 2010
- [4]GATELY, E. *Neural Networks for Financial Forecasting*. John Wiley & Sons, Inc, United States, 1996.
- [5]KUPFER, D.; HASENCLEVER, L. *Economia Industrial*. Rio de Janeiro: Elsevier, 2002.
- [6]ÖBERG, C.; HOLTSTRÖM, J. Are mergers and acquisitions contagious? *Journal of Business Research*, Amsterdam, v. 59, p. 1267–1275, 2006.
- [7]VIEIRA, A.A; Cavalcanti, V.G. Avaliação de insolvência no sistema bancário: uma aplicação para o caso brasileiro. *Revista Desenbahia* nº 13 / set. 2010

# Avaliação de um sistema de medição via índice R&R pelo método da Análise de Variância

Eduardo Campana Barbosa<sup>1</sup>

Bruno Carlos Alves Pinheiro<sup>2</sup>

Sérgio Célio da Silva Lima<sup>3</sup>

Carlos Henrique Osório Silva<sup>4</sup>

Lucas Carvalho Gonçalves<sup>5</sup>

**Resumo:** Serão apresentados os resultados de um estudo sobre a análise do parâmetro R&R em um processo de medição da densidade após secagem de corpos-de-prova de revestimento cerâmico para piso. O parâmetro R&R foi obtido estimando-se a variabilidade associada à repetitividade, a reprodutibilidade e a variação total do processo, por meio do método da Análise de Variância. Os resultados mostraram que o processo de medição não é aceitável e precisa ser melhorado. Concluiu-se também que a Repetitividade é a componente com maior influência para o desajuste do processo.

**Palavras-chave:** Medição; Repetitividade; Reprodutibilidade, Controle de Qualidade.

## 1 Introdução

Um processo de medição pode sofrer a influência de vários fatores, como instabilidades ambientais, diferenças entre os operadores etc., o que pode conduzir a uma incerteza associada ao mesmo. A confiabilidade das medições deve ser analisada com o auxílio de ferramentas estatísticas, onde os resultados são submetidos a critérios de aceitação, comparação, investigação ou avaliação de um sistema de medição [1]. O parâmetro R&R é um índice utilizado para verificar a adequabilidade de um sistema de medição e pode-se determiná-lo por: (a) método da amplitude, (b) método da média e amplitude e (c) método da Análise de Variância (ANOVA) com **efeitos aleatórios**. Abordaremos a determinação do parâmetro R&R pelo método da ANOVA, o qual é descrito a seguir.

Em estudos do parâmetro R&R a variabilidade total ( $\sigma_{Total}^2$ ) dos valores de uma variável aleatória Y (característica da qualidade) é decomposta conforme em (1):

---

<sup>1</sup>DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: *duducampana@hotmail.com*

<sup>2</sup>DPP-UEMG: Universidade Estadual de Minas Gerais, Ubá-MG. Email: *Bruno.pinheiro@uemgubá.edu.br*

<sup>3</sup>DEP-FIC/UNIS: Faculdades Integradas de Cataguases, Cataguases-MG. Email: *sergiolima\_500@yahoo.com.br*

<sup>4</sup>DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: *chos@ufv.br*

<sup>5</sup>CEFET-MG: Centro Federal de Educação Tecnológica de Minas Gerais. Email: *lucas.encautcefemg@gmail.com*

$$\sigma_{Total}^2 = \sigma_{Peça}^2 + \sigma_{Medição}^2 \quad (1)$$

Em que  $\sigma_{Peça}^2$  é a variabilidade inerente às médias das peças avaliadas e  $\sigma_{Medição}^2$  é a variabilidade associada aos erros de medição. Temos que  $\sigma_{Medição}^2$  pode ser expresso como em (2):

$$\sigma_{Medição}^2 = \sigma_{Repetitividade}^2 + \sigma_{Reprodutividade}^2 \quad (2)$$

Segundo [2], a Repetitividade refere-se à variância associada aos valores de Y, obtidos por meio de medições sucessivas de uma mesma peça, mensuradas por um mesmo operador e instrumento. A Reprodutividade refere-se à variabilidade das médias obtidas por diferentes operadores, ao realizar a medição repetidas vezes com um mesmo instrumento. Por isso o nome R&R.

Para avaliar um sistema de medição é necessário obter as estimativas das variâncias citadas. A Tabela 1 apresenta um esquema geral pelo método da ANOVA com efeitos aleatórios. Tem-se que **P** é o Fator Peça, **O** é o Fator Operador e **PO** é a interação entre o Fator Peça e o Fator Operador. Considere um experimento balanceado e instalado sob o DIC, em que **n** peças e **p** operadores são selecionados aleatoriamente, dado ainda que cada operador mede cada peça **r** vezes.

Tabela 1 – Esquema da Análise de Variância (ANOVA) com interação e efeitos aleatórios

<i>F.V</i>	<i>G.L</i>	<i>S.Q</i>	<i>Q.M</i>	<i>Teste F</i>	<i>E(QM)</i>
<i>P</i>	$(n - 1)$	<i>SQP</i>	$SQP/(n - 1)$	$QMP/QM(PO)$	$\sigma_R^2 + r\sigma_{(PO)}^2 + pr\sigma_P^2$
<i>O</i>	$(p - 1)$	<i>SQO</i>	$SQO/(p - 1)$	$QMO/QM(PO)$	$\sigma_R^2 + r\sigma_{(PO)}^2 + nr\sigma_O^2$
<i>PO</i>	$(n - 1)(p - 1)$	<i>SQ(PO)</i>	$SQ(PO)/(n - 1)(p - 1)$	$QM(PO)/QMR$	$\sigma_R^2 + r\sigma_{(PO)}^2$
<i>Resíduo</i>	$np(r - 1)$	<i>SQR</i>	$SQR/np(r - 1)$		$\sigma_R^2$
<i>Total</i>	$npr - 1$				

[1] e [3] explicam que em estudos do parâmetro R&R, tem-se que:

$$\sigma_{Repetitividade}^2 = \sigma_R^2 \quad (3)$$

$$\sigma_{Reprodutividade}^2 = \sigma_O^2 + \sigma_{(PO)}^2 \quad (4)$$

As estimativas de cada componente da variância podem ser obtidas utilizando-se o método dos momentos, que consiste em igualar a Esperança do Quadrado Médio de cada termo a seu respectivo Quadrado Médio. Com uma simples modelagem matemática é possível encontrar os resultados demonstrados (5), (6), (7) e (8).

$$\hat{\sigma}_R^2 = QMR \quad (5)$$

$$\hat{\sigma}_{(PO)}^2 = \frac{QM(PO) - QMR}{r} \quad (6)$$

$$\hat{\sigma}_P^2 = \frac{QMP - QM(PO)}{pr} \quad (7)$$

$$\hat{\sigma}_O^2 = \frac{QMO - QM(PO)}{nr} \quad (8)$$

Posteriormente o valor do índice R&R deve ser calculado conforme as equações abaixo:

$$R\&R = 6 \cdot \hat{\sigma}_{medição} \quad (9)$$

$$R\&R (\%) = 100 \cdot \frac{R\&R}{6 \cdot \hat{\sigma}_{Total}} = 100 \cdot \frac{\hat{\sigma}_{medição}}{\hat{\sigma}_{Total}} \quad (10)$$

$$\hat{\sigma}_{medição} = \sqrt{\hat{\sigma}_{Repetitividade}^2 + \hat{\sigma}_{Reprodutividade}^2} = \sqrt{\hat{\sigma}_{Resíduo}^2 + (\hat{\sigma}_O^2 + \hat{\sigma}_{(PO)}^2)} \quad (11)$$

$$\hat{\sigma}_{Total} = \sqrt{\hat{\sigma}_{Peça}^2 + \hat{\sigma}_{Medição}^2} = \sqrt{\hat{\sigma}_P^2 + [\hat{\sigma}_{Resíduo}^2 + (\hat{\sigma}_O^2 + \hat{\sigma}_{(PO)}^2)]} \quad (12)$$

Destaca-se que se algum componente da variância apresentar valor negativo adota-se o valor zero (0). Caso a interação não seja significativa (valor-p >  $\alpha$ ) deve-se refazer a ANOVA sem este termo e os componentes  $\hat{\sigma}_P^2$  e  $\hat{\sigma}_O^2$  serão obtidos subtraindo-se o *QMP* e *QMO* do *QMR* ao invés de *QM(PO)*, e dividindo por *pr* e *nr*, respectivamente. Após a obtenção do índice R&R (%), os critérios para avaliação de um processo de medição são apresentados na Tabela 2:

Tabela 2 – Critérios para avaliação de um processo de medição via parâmetro R&R. Fonte: [1].

R&R (%)	Conclusão
R&R < 10%	Processo de Medição Aceitável.
10% ≤ R&R ≤ 30%	Processo de Medição considerado adequado, dependendo da importância do estudo.
R&R > 30%	Processo de Medição é Inaceitável.

## 2 Materiais e Métodos

Neste trabalho, a característica da qualidade (Y) foi representada pelas medidas de massa específica aparente após secagem ( $\rho_s$ ) de corpos-de-prova de revestimento cerâmico para piso. As amostras foram obtidas por prensagem uniaxial com pressão de compactação de 50 Mpa e posteriormente submetidas a um processo de secagem em estufa de laboratório a 110°C por 24 h. Em seguida, mediram-se as dimensões (comprimento, largura e espessura) de cada amostra para a determinação do volume ( $V_s$ ) e também sua massa ( $M_s$ ). Os instrumentos utilizados foram: a) paquímetro da marca Mitutoyo (0,01 mm) e, b) balança digital ( $\pm 0,01$  g). A massa específica aparente após secagem foi determinada pela fórmula  $\rho_s = \frac{M_s}{V_s}$ .

Para o processo de medição (planejamento experimental) foram selecionados aleatoriamente dois operadores e dez corpos-de-prova. Cada amostra foi medida repetidamente cinco vezes por cada operador. Utilizou-se a função *aov* do *software* livre R para o desenvolvimento das análises estatísticas. As Tabelas 3 e 4 apresentam, respectivamente, para os operadores 1 e 2, as medições dos valores de massa específica aparente após secagem (densidade a seco) para os 10 corpos-de-prova.

Tabela 3 – Medições realizadas pelo Operador 1

Medição/peças	1	2	3	4	5	6	7	8	9	10
1	1,85	1,88	1,87	1,89	1,87	1,9	1,89	1,87	1,9	1,89
2	1,86	1,91	1,89	1,85	1,86	1,87	1,92	1,87	1,87	1,93
3	1,87	1,88	1,88	1,86	1,89	1,88	1,91	1,90	1,91	1,90
4	1,82	1,87	1,88	1,87	1,87	1,92	1,88	1,89	1,88	1,86
5	1,86	1,89	1,87	1,87	1,86	1,87	1,91	1,87	1,88	1,88

Tabela 4 – Medições realizadas pelo Operador 2

Medição/peças	1	2	3	4	5	6	7	8	9	10
1	1,88	1,91	1,90	1,90	1,87	1,88	1,89	1,91	1,89	1,89
2	1,90	1,88	1,90	1,91	1,90	1,89	1,89	1,90	1,91	1,87
3	1,92	1,90	1,91	1,90	1,90	1,91	1,89	1,91	1,90	1,90
4	1,90	1,91	1,87	1,90	1,89	1,90	1,89	1,86	1,90	1,91
5	1,86	1,89	1,89	1,89	1,88	1,91	1,91	1,89	1,88	1,89

### 3 Resultados e Discussão

Na Tabela 5 apresenta os resultados da ANOVA.

Tabela 5 – Resultados da ANOVA sem a interação Peça x Operador

F.V	G.L	S.Q	Q.M	Teste F	valor-p
Peça	9	0,005285	0,0005872	2,1117	0,0365*
Operador	1	0,005041	0,0050410	18,128	0,0000*
Resíduo (Repetitividade)	89	0,024749	0,0002781		
Total	99	0,035075			

Os Fatores Peça e Operador foram significativos a 5% de probabilidade (valor-p < 0,05), demonstrando que existe diferença estatística entre as peças e entre a medição dos operadores. Os resíduos da ANOVA foram normais segundo o teste de Shapiro-Wilk (valor-p = 0,6924), homogêneos pelo teste de Bartlett (valor-p = 0,3436) e aleatórios segundo o runs.test (valor-p = 0,6877), ver detalhes na Figura 1. Pode-se observar que não se tem a presença da interação como F.V na Tabela 5, visto que esta não foi significativa (valor-p > 0,05). Logo, a estimativa do parâmetro R&R é dada por:

$$\hat{\sigma}_P^2 = 0,0000309 \quad \hat{\sigma}_O^2 = 0,00009526$$

$$\hat{\sigma}_{\text{medição}} = \sqrt{0,0002781 + (0,00009526 + 0)} = 0,01932$$

$$\hat{\sigma}_{\text{Total}} = \sqrt{0,0000309 + [0,0002781 + (0,00009526 + 0)]} = 0,0201064$$

$$R\&R (\%) = 100 * \frac{0,01932}{0,02010} = 96,10\%$$

Como o índice R&R (%) apresenta valor superior a 30% pode-se concluir que o sistema de medição não está adequado. Na Tabela 6 encontram-se indicados as estimativas dos percentuais de contribuição ou o percentual da variabilidade total explicada por cada componente da variância.

Tabela 6 – Percentual da Variância Total explicado por cada componente

	Peça	Operador	Repetitividade	Reprodutividade
$\widehat{\sigma}_{\text{componente}}^2$	0,0000309	0,00009526	0,0002781	0,00009526
$\widehat{\sigma}_{\text{Total}}^2$	0,0004043	0,0004043	0,0004043	0,0004043
% Variância	7,64%	23,56%	68,79%	23,56%

Nota-se que o maior percentual da variabilidade total do sistema de medição pode ser explicado pela componente que representa a Repetitividade (68,79%). Logo, conclui-se que o maior problema do sistema está associado à qualidade das medições efetuadas pelos operadores. Os possíveis

fatores causadores de tal problema poderiam ser a falta de habilidade, a baixa acuidade visual e/ou ainda a falta de cuidado ao efetuar a medição.

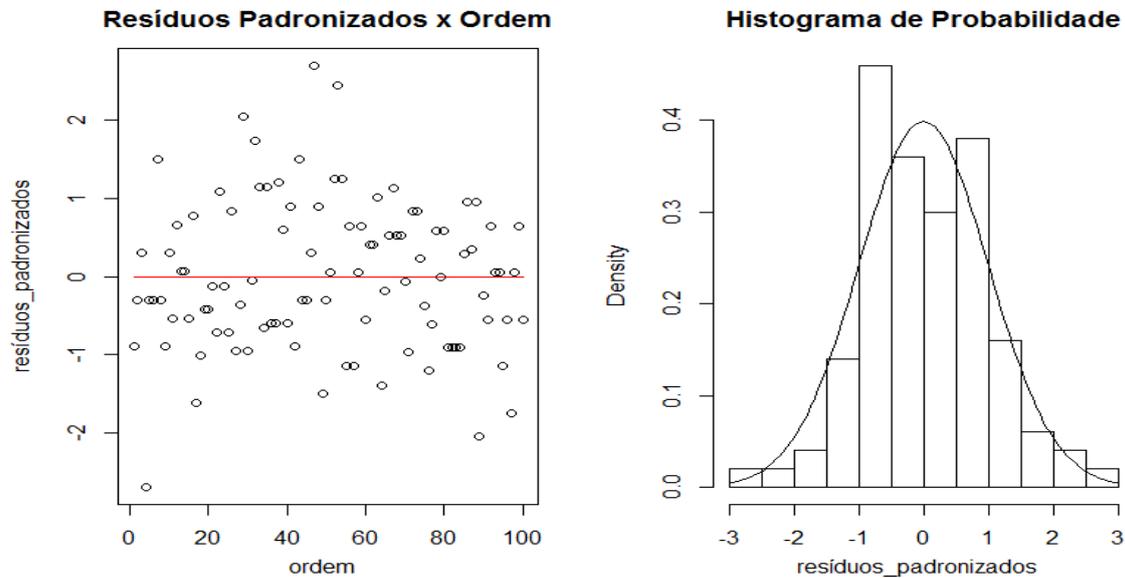


Figura 1 – Gráfico dos Resíduos Padronizados da ANOVA.

## 4 Conclusões

O índice R&R estimado via ANOVA é de fácil execução, mostrando-se útil para avaliar o sistema de medição em questão, evidenciando o seu desajuste ( $R\&R > 30\%$ ) e indicando que a Repetitividade entre operadores é o componente com maior influência para este desajuste. Sugere-se que treinamentos sejam oferecidos aos operadores, no intuito de capacitá-los para que realizem de forma correta e eficiente tais medições, no intuito de adequar o sistema. Para estudos futuros, sugere-se que um novo estudo R&R seja desenvolvido após a realização de tais treinamentos, no intuito de avaliar se houve melhoria no índice R&R (%).

## 5 Referências Bibliográficas

- [1] JUNIOR, A.A.G; SOUSA, A.R. *Fundamentos de metrologia científica e industrial*. 1<sup>a</sup>. ed. Barueri - SP: Manole, 2008.
- [2] RIBEIRO JÚNIOR, J. I. *Métodos Estatísticos Aplicados à Melhoria da Qualidade*. UFV, 2012.
- [3] MONTGOMERY, D.C. *Introdução ao controle estatístico da qualidade*. 4<sup>a</sup>. ed. Rio de Janeiro: LTC, 2004.

# Índice de massa corporal de adolescentes e atividade física: um estudo de caso com diferentes abordagens de análise

Helen Maria Pedrosa de Oliveira<sup>1 3</sup>

Eric Batista Ferreira<sup>2 3</sup>

**Resumo:** O número de pessoas acima do peso vem crescendo muito nos últimos anos e parte disso deve à má alimentação da população aliada à falta da prática de exercícios físicos regulares. O trabalho de Costa (2010) apresenta um experimento que relaciona o Índice de Massa Corporal (IMC) com a prática de atividade física de adolescentes em Portugal. O objetivo deste trabalho é apresentar diferentes abordagens analíticas deste mesmo experimento. Tais reflexões são fundamentais no planejamento e consequente análise de experimentos similares. Há indícios para crer que o IMC pode ser diferente entre homens e mulheres e, além disso, como não é foco principal a comparação dos gêneros, talvez seja mais aconselhado analisá-los em dois experimentos distintos. Ainda, a prática de atividade física não interferiu no IMC dos adolescentes.

**Palavras-chave:** Índice de Massa Corporal, Análise de variância, Modelo aleatório, Componentes de variância.

## 1 Introdução

Atualmente, o número de pessoas que estão acima do peso vem crescendo consideravelmente. Em geral, não é difícil reconhecer a obesidade ou até mesmo o sobrepeso, mas o diagnóstico correto requer que se identifiquem os níveis de risco, o que, frequentemente, necessita de algumas formas de quantificação. O Ministério da Saúde considera “acima do peso” as pessoas com um Índice de Massa Corporal (IMC) maior que 25. O IMC é calculado dividindo-se o peso pela altura ao quadrado. Acima de 30, a pessoa ser considerada “obesa”. (PASSARINHO, 2012)

Segundo o Ministério da Saúde, quase metade da população brasileira sofre com este problema. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), o sobrepeso atinge mais de 30% das crianças entre 5 e 9 anos de idade, cerca de 20% da população entre 10 e 19 anos e nada menos que 48% das mulheres e 50,1% dos homens acima de 20 anos. Mantido o ritmo atual de crescimento do número de pessoas acima do peso, em dez anos 30% da população

---

<sup>1</sup>Mestranda - UNIFAL-MG: Universidade Federal de Alfenas. Email: [helen.pedrosa@yahoo.com.br](mailto:helen.pedrosa@yahoo.com.br)

<sup>2</sup>ICEx - UNIFAL-MG: Universidade Federal de Alfenas. Email: [eric.ferreira@unifal-mg.edu.br](mailto:eric.ferreira@unifal-mg.edu.br)

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

será obesa, resultando em um sério problema de saúde pública, já que o sobrepeso impulsiona o aparecimento de doenças crônicas (SOARES; RITTO, 2010).

Apesar de a obesidade estar relacionada a fatores genéticos, há importante influência do sedentarismo e de padrões alimentares inadequados no aumento dos índices brasileiros. A prática regular de exercícios físicos está longe de fazer parte dos hábitos dos mesmos. Soares e Ritto (2010) afirmam que uma pesquisa de 2008 do IBGE mostrou que apenas 10,2% da população com 14 anos ou mais tem alguma atividade física regular.

O tratamento da obesidade é complexo e multidisciplinar. Não existe nenhum tratamento farmacológico em longo prazo que não envolva a mudança de estilo de vida. A obesidade é um desafio mundial, pelo que representa de redução na expectativa de vida e nos custos dos serviços de saúde (ABESO, 2009).

## 2 Objetivo

Discutir diferentes abordagens analíticas de um experimento desbalanceado cujo foco é a relação existente entre o índice de massa corporal e a atividade física de adolescentes.

## 3 Metodologia

Este trabalho tomou como base os dados que podem ser encontrados no trabalho de Costa (2010). Os dados são oriundos de uma pesquisa feita entre 2008 e 2009 com 299 alunos de 15 à 19 anos de quatro escolas de Ensino Médio de Portugal: E.S. Homem Cristo, E.S. José Estêvão, E.S. Dr. Jaime Magalhães Lima e E.S. Mário Sacramento. Os alunos foram divididos em grupos diferentes, de acordo com o gênero (feminino e masculino), a prática de atividade física em anos anteriores (sim e não) e a prática de atividade física atualmente (sim e não), que resultou na combinação de um fatorial triplo com dois níveis: **FSS**: mulheres que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa e atualmente; **FSN**: mulheres que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa, mas não atualmente; **FNS**: mulheres que exercitaram a prática de exercícios físicos atualmente, mas não em anos anteriores; **FNN**: mulheres que nunca exercitaram a prática de exercícios físicos; **MSS**: homens que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa e atualmente; **MSN**: homens que exercitaram a prática de exercícios físicos em anos anteriores à pesquisa, mas não atualmente; **MNS**: homens que exercitaram a prática de exercícios físicos atualmente, mas não em anos anteriores; **MNN**: homens que nunca exercitaram a prática de exercícios físicos.

Após a divisão dos grupos, foi feita a medição do Índice de Massa Corporal (IMC), com o

objetivo de avaliar o peso corporal dos jovens. Este índice é definido pelo quociente entre o peso (em quilogramas) e o quadrado da altura (em metros). O valor assim obtido estabelece o diagnóstico da obesidade e caracteriza também os riscos associados.

Os dados de Costa (2010), que serão utilizados neste trabalho apresentam 10,5% de adolescentes abaixo do peso normal, quase 80% de pessoas no peso saudável, quase 10% de pessoas com sobrepeso e 1% de adolescentes com obesidade de grau I.

Todas as análises e figuras foram feitas no software R (R CORE TEAM, 2014). Em particular, as análises de variância foram feitas por meio do pacote `ExpDes.pt` (FERREIRA; CAVALCANTI; NOGUEIRA, 2013).

## 4 Resultados e Discussão

Como primeira atitude perante os dados, foi realizada uma análise exploratória. A Tabela 1 explicita o desbalanceamento previamente conhecido. Foi computado o índice de desbalanceamento, utilizado por Martins Jr., Ferreira e Ramos (2013),  $\delta = \frac{\max(J_i)}{\min(J_i)}$ , em que  $J_i$ , com  $i = 1, \dots, 8$  é o número de repetições do  $i$ -ésimo tratamento. Para a presente amostra,  $\delta = 6,44$ ; o que sugere um grande desbalanceamento, uma vez que esses mesmos autores, em seu estudo de simulação, trabalharam com índices variando entre 1,17 e 8.

Em seguida, foram feitos box-plots para os tratamentos a fim de detectar possíveis valores discrepantes. A Figura 1(a) sugere a presença de 1, 3 e 1 *outliers*, respectivamente nos tratamentos FNS, FSS e MSS. Ressalta-se que os três tratamentos que apresentaram valores discrepantes receberam resposta “sim” à pergunta sobre a prática de atividade física atualmente. Após a remoção dos valores discrepantes, a Figura 1(b) ainda apresenta indícios de que pode haver diferença entre as medianas dos tratamentos e que as variâncias podem não ser homogêneas.

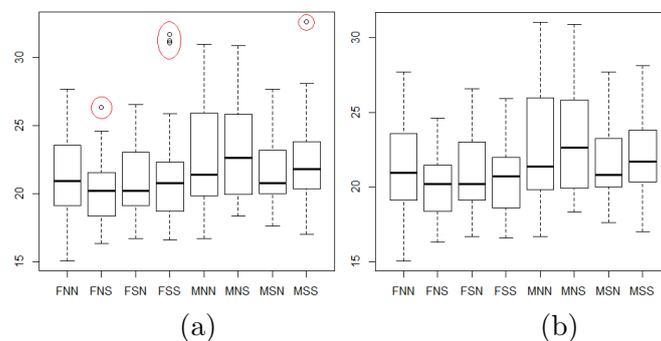


Figura 1: *Box-plots* dos dados de IMC para cada tratamento, antes (a) e após (b) a remoção das observações discrepantes.

Executaram-se testes de normalidade para todos os tratamentos por meio do teste de Shapiro-Wilk à 5% de significância, mostrando que para nenhum dos tratamentos deve-se rejeitar a

hipótese nula de normalidade (Tabela 1).

Tabela 1: Teste de normalidade Shapiro-Wilk, a 5% de significância, por tratamento.

<b>Tratamentos</b>	FNN	FNS	FSN	FSS	MNN	MNS	MSN	MSS
<b>Repetições</b>	43	58	17	58	13	20	9	81
<b>Valor-p</b>	0,3662	0,0928	0,5161	0,0578	0,6102	0,1466	0,6306	0,1698

Em seguida procedeu-se a análise de variância considerando um modelo fixo em delineamento inteiramente casualizado (DIC). Com seus resíduos, foram verificadas as pressuposições de normalidade, homogeneidade de variâncias e independência, utilizando-se os testes de Shapiro-Wilk ( $p=0,015$ ), Bartlett ( $p=0,4180$ ) e Durbin-Watson ( $p=0,1456$ ), respectivamente, todos a 5% de significância. Nota-se que, embora os tratamentos sejam individualmente normais, os resíduos do modelo fixo não podem ser considerados como tal. Por outro lado, não se pode dizer as pressuposições de homoscedasticidade e independência foram violadas.

O teste F da ANAVA acusa a diferença entre médias ( $p=0,0013$ ), tendo o experimento um coeficiente de variação baixo (13,18%). Simultaneamente foi realizado o teste não paramétrico de Kruskal-Wallis. Este também apontou que há diferença entre os tratamentos ( $p=0,0056$ ).

Então partiu-se para os testes de comparações múltiplas de médias. Foram realizados os testes de Tukey, Student-Newman-Keuls (SNK), Bonferroni (Bonf), Bootstrap, Scott-Knott (SK), t de Student (LSD), Duncan e Calinski & Corsten bootstrap (CCF), sempre considerando-se 5% de significância. Todos concordaram com o teste F, exceto o teste de Bootstrap de Ramos e Ferreira (2009), que se mostrou conservador.

A ordenação das médias, mostra uma clara divisão entre mulheres e homens, com os maiores IMCs sendo os masculinos, como mostra a Tabela 2.

Tabela 2: Média de cada um dos tratamentos.

<b>Tratamentos</b>	MNS	MNN	MSS	MSN	FSS	FNN	FSN	FNS
<b>Média</b>	22,90	22,53	22,12	21,73	21,41	21,22	20,97	20,29

Com isso, uma outra abordagem - talvez ainda mais útil - seja a partição da fonte de variação tratamentos em três novas fontes: “gênero”, “atividade física atual” e “atividade física anterior”. Ainda considerando um modelo fixo em DIC, mas dessa vez em esquema fatorial triplo, uma nova análise da variância foi feita.

Com esse tipo de abordagem conseguimos observar os efeitos simples. De acordo com o teste F, à 5% de significância temos que os fatores prática de exercícios físicos em anos anteriores e atualmente também foram não significativos. Por sua vez, o fator “gênero” se mostrou significativo, apontando que no gênero feminino a Índice de Massa Corporal é mais baixo (20,95 contra 22,26 do masculino), suspeita já existente desde a abordagem em DIC simples.

## 5 Conclusões

Neste trabalho foi possível discutir diversas abordagens de análise de experimentos desbalanceados que tratam da comparação de índices de massa corporal (IMC). Tais reflexões são fundamentais no planejamento e consequente análise de experimentos similares.

Além disso, pode-se observar uma possível discrepância entre IMCs de adolescentes dos gêneros masculino e feminino. Mesmo sendo uma métrica que pondera o peso pelo quadrado da altura, o IMC talvez não seja capaz de deixar homens e mulheres comparáveis. Por sua vez, a prática atual ou pregressa de atividade física não pareceu interferir no IMC de adolescentes dessa região do Portugal.

## Referências

- [1] ABESO - Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica. *Diretrizes brasileiras de obesidade 2009/2010*, 3ª ed. - Itapevi, SP: AC Farmacêutica, 2009.
- [2] FERREIRA, E. B., CAVALCANTI, P. P., NOGUEIRA, D. A. *ExpDes.pt: Experimental Designs* pacakge (Portuguese). R package version 1.1.2. 2013.
- [3] COSTA, C. P. S. *Um estudo sobre a robustez do teste-F*. Dissertação (Mestrado em Matemática e Aplicações) - Universidade de Aveiro, Portugal. 2010. 114 p.
- [4] MARTINS JR., J. M.; FERREIRA, E. B., RAMOS, P. S. Alternatives for correcting the Tukey's statistics for unbalanced experiments. *Sigmae*, Alfenas, v.2, n.2, p. 21-42. 2013.
- [5] PASSARINHO, N. *Quase metade da população está acima do peso, diz Saúde*. 2010. Acesso em: 23 jul 2014. Disponível em: <http://g1.globo.com/ciencia-e-saude/noticia/2012/04/quase-metade-da-populacao-esta-acima-do-peso-diz-saude.html>.
- [6] PIMENTEL-GOMES, F. *Curso de estatística experimental* 15.ed. Piracicaba:FEALQ,2009.
- [7] R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2014.
- [8] RAMOS, P. S., FERREIRA, D. F. Agrupamento de medias via bootstrap de populacoes normais e nao-normais, *Revista Ceres*, v.56, p.140-149, 2009.
- [9] SOARES, L.; RITTO, C. *Pesquisa do IBGE confirma que obesidade é epidemia no Brasil*. 2010. Acesso em: 22 jul 2014. Disponível em: <http://veja.abril.com.br/noticia/saude/pesquisa-do-ibge-mostra-que-obesidade-e-epidemia-no-brasil>.

# Previsão das estratégias de Governança Corporativa: um estudo econométrico no período de 2002 a 2013

Jennifer Elis Pires<sup>1 3</sup>

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

**Resumo:** A literatura brasileira vem abordando diversas linhas de trabalho que estudam temáticas associadas à maximização do desempenho organizacional. Dentro desta discussão o tema da governança corporativa tem recebido atenção de muitos pesquisadores, uma vez que tem sido visto como forma da empresa atingir melhores desempenhos financeiros e econômicos. Neste contexto, o objetivo deste trabalho é desenvolver um modelo de previsão de governança corporativa, com base em indicadores financeiros e econômicos, para instituições não bancárias brasileiras utilizando regressão logística. Selecionaram-se para este estudo 33 empresas, com ações negociadas na BM&FBovespa. As empresas foram selecionadas entre as 100 ações mais negociadas no ano de 2013. A amostra utilizada abrange o período de 2002 a 2013, os dados são trimestrais e totalizam 1476 observações. Os resultados demonstram que a regressão logística foi capaz de classificar corretamente 81,1% das empresas que fazem parte dos níveis de governança corporativa. A previsão do modelo para as empresas que possuem governança corporativa foi de 97,6%, enquanto para aquelas que não possuem governança corporativa foi de 18,20%. Assim os resultados encontrados confirmam que indicadores como liquidez geral, liquidez corrente, participação de terceiro, imobilização do patrimônio líquido e retorno sobre o ativo total determinam o mecanismo de Governança Corporativa.

**Palavras-chave:** *Governança Corporativa, Logit, finanças.*

## 1.Introdução

Uma tendência dos últimos anos é a instituição de práticas de boa governança corporativa nas empresas, esse é atualmente motivo de discussão em diversas áreas como administração de empresas, direito, economia e finanças. A governança corporativa pode ser definida como um conjunto de valores e normas que visa diminuir conflitos de interesse, maximizar a riqueza dos acionistas, buscando sempre levar em conta os interesses das minorias, mas atentando a otimizar o desempenho de uma companhia, promovendo o desenvolvimento e o crescimento econômico (IQUIPAZA et al, 2007).

---

<sup>1</sup> ICESA – UNIFAL: Universidade Federal de Alfenas, Varginha-MG. Email: *jenni.elis@hotmail.com*

<sup>2</sup> ICESA – UNIFAL: Universidade Federal de Alfenas Varginha-MG. Email: *gabriel.pessanha@unifal-mg.edu.br*

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

O mercado financeiro tem dado mais atenção à governança em função das mudanças ocorridas no cenário mundial e dos escândalos corporativos dos últimos anos, principalmente no mercado norte-americano, que é reconhecido como o principal fornecedor de capital do mundo. Além disso, os processos de privatização, fusões e aquisições, e o crescimento dos fundos de pensão também contribuíram para a preocupação com os estudos da governança das corporações (ALMEIDA et al, 2010). Todo esse processo tem impactado e promovido mudanças nas legislações dos diversos países, os quais têm buscado estabelecer regras mais claras e transparentes como forma de garantir maior proteção aos investidores e criar um ambiente favorável para o alinhamento de interesses.

No Brasil, em 2000, a BM&F BOVESPA criou o Nível 1, o Nível 2 e o Novo Mercado, denominados de Níveis Diferenciados de Governança Corporativa, ou NDGC. O objetivo geral é fortalecer o mercado acionário, proporcionar aos investidores um ambiente de maior transparência, proteger os acionistas minoritários e, conseqüentemente, favorecer as organizações. Ao mesmo tempo em que a adoção às práticas diferenciadas de governança corporativa faz aumentar a demanda pelas ações da empresa, ela também faz com que os investidores ocorram em menor risco, tornando o mercado acionário uma melhor alternativa de investimento. Isso, por sua vez, afeta a economia nacional, com uma maior canalização da poupança nas empresas (BM&F BOVESPA, 2013).

## 2. Objetivo

Devido à importância que a governança corporativa vem apresentando perante a economia brasileira, aos investidores e as empresas o objetivo deste trabalho é desenvolver um modelo de previsão de governança corporativa, com base em indicadores financeiros e econômicos, para instituições não bancárias brasileiras utilizando regressão logística.

## 3. Metodologia

Selecionou-se para este estudo 33 empresas não bancárias, com ações negociadas na Bolsa de Valores de São Paulo (Bovespa), das quais 26 apresentam práticas de governança. A amostra utilizada abrange o período de 2002 a 2013, na pesquisa os dados são trimestrais e consiste de 1476 observações. Os dados utilizados nas análises estatísticas foram obtidos através do banco de dados Economatica. O *software* utilizado para o tratamento dos dados e ajuste do modelo de regressão logística foi o *gretl* versão 1.9.90.

Na equação 1 é demonstrado o modelo de regressão logística utilizado nesta pesquisa, no qual se utilizou das variáveis descritas na Tabela 1. Assim:

$$GC_i = \beta_0 + \beta_1 LG_i + \beta_2 LC_i + \beta_3 LS_i + \beta_4 PT_i + \beta_5 CE_i + \beta_6 IPL_i + \beta_7 ROE_i + \beta_8 ROA_i + \beta_9 ML_i + \beta_{10} TAM_i + \varepsilon \quad (1)$$

## 4. Análise dos Resultados

Na Tabela 1 são apresentados dois modelos de previsão para a governança corporativa, no Modelo 1 utilizou-se todos os indicadores, tanto financeiro quando econômico, contudo é possível

observar que algumas variáveis não apresentaram significância para o modelo. Assim, para mitigar o efeito de potenciais variáveis não relacionadas à escolha, dado o tamanho da amostra, foi empregado o procedimento *stepwise*<sup>4</sup>. Através desse método conseguiu-se elaborar um melhor ajuste demonstrado no Modelo 2.

Tabela 1 – Estimação do Modelo para Governança Corporativa

VARIÁVEIS	MODELO 1		MODELO 2	
	Coeficiente	P - valor	Coeficiente	P – valor
<b>Constante</b>	-1,06296	0,6443	1,09574	9,21e-09 ***
<b>LG</b>	-0,0576517	0,0389 **	-0,0580527	0,0294 **
<b>LC</b>	0,710402	0,0034 ***	0,697584	6,42e-09 ***
<b>LS</b>	-0,0727111	0,8170	-	-
<b>PT</b>	-0,0411920	0,0252 **	-0,0374295	0,0124 **
<b>CE</b>	-0,0788422	0,8641	-	-
<b>IPL</b>	0,0948420	0,0426 **	0,110020	0,0125 **
<b>ROE</b>	-0,621953	0,1949	-	-
<b>ROA</b>	-13,6459	6,93e-012 ***	-16,1268	7,51e-030 ***
<b>ML</b>	-0,823134	0,1122	-	-
<b>TAM</b>	0,856698	0,2990	-	-
	<b>PREVISÃO</b>		<b>PREVISÃO</b>	
	80,9%		81,1%	
*** Significativo a 1% ; ** Significativo a 5% ; * Significativo a 10%.				

A natureza das variáveis independentes e seus efeitos no ajuste do modelo produziram coeficientes de regressão logística que devem ser interpretados com cautela. Variáveis com coeficiente logístico estimado negativo indicam que a categoria focalizada, em relação à referência, está associada com a diminuição na probabilidade de possuir governança corporativa. Enquanto variáveis com coeficiente logístico estimado positivo indicam que a categoria estudada, em relação a referência, está associada com o aumento da probabilidade de possuir governança corporativa.

Desse modo, o modelo 2 apresenta que o coeficiente de liquidez geral (LG) é negativo, indicando que quando maior for essa liquidez menor será a probabilidade da empresa possuir

<sup>4</sup> O procedimento constrói iterativamente uma sequência de modelos de regressão pela adição ou remoção de variáveis em cada etapa.

governança corporativa, contudo o coeficiente é bem pequeno, de apenas -0,05, essa relação inversa entre governança e liquidez geral pode ter ocorrido devido ao fato que esse índice leva em consideração direitos e deveres de longo prazo, podendo assim resultar num valor negativo.

Enquanto a liquidez corrente (LC) demonstrou um coeficiente de aproximadamente 0,69, ou seja, quanto maior a liquidez corrente maior será a probabilidade de ter governança corporativa, o que conduz com a teoria apresentada, de modo que as empresas que possuem governança corporativa estão expostas a menores riscos de liquidez, resultando assim num salto positivo da liquidez.

O coeficiente da participação de terceiros (PT) foi de aproximadamente -0,03, ao nível de 95% de confiança, demonstrando que o aumento da participação de terceiros diminui a probabilidade de ter governança corporativa. Esse evento é coerente, pois quando a empresa passa a possuir governança corporativa seu risco de liquidez diminui, de modo que passa a atrair mais fundos para o mercado de capitais, inibindo assim a necessidade de obter capital de terceiro porque seu capital próprio já é suficiente.

A imobilização do patrimônio líquido (IPL) apresentou um coeficiente de 0,11, ao nível de 5% de significância, assim quanto mais aplicações do patrimônio líquido forem feitas no ativo permanente maior será a probabilidade da empresa possuir governança corporativa. Isto pode ser explicado pelo porte das empresas estudadas e pela representatividade dos seus ativos imobilizados.

O retorno sobre o ativo total (ROA) obteve um resultado inesperado, no qual seu coeficiente, ao nível de 99% de confiança, foi de -16,12 esse resultado não era esperado, pois a teoria vem demonstrando que empresas que possuem governança corporativa tentem a aumentar seus retornos para passar maior confiabilidade para seus clientes e acionistas. Segundo relatórios técnicos da BM&FBovespa (2013) quando uma empresa passa a fazer parte de um dos níveis de governança corporativa a sua demanda por suas ações passar a ser maior, assim como custo de capital diminui, esse fato implicaria num resultado melhor para empresa, contudo não foi o que ocorreu com as empresas investigadas.

Tabela 2 – Tabela de Classificação do Modelo de Regressão Logística

		Real		Total
		ÑGC	GC	
Previsto	ÑGC	56	251	307
	GC	28	1141	1169
	Total	84	1392	1476

**Acerto = 81,1%**

De acordo com os resultados mostrados na Tabela 2, o modelo geral teve aproximadamente 81% de acerto em sua estimativa sobre as empresas que fazem parte dos níveis diferenciados de governança corporativa da Bovespa, e consequentemente possuem governança corporativa. O nível de

previsão do modelo para as empresas que possuem governança corporativa foi de 97,6%, enquanto o nível de previsão para aquelas que não possuem governança corporativa foi de 18,20%.

É possível observar que das 307 variáveis que o modelo previu não possuir governança corporativa, 56 de fato não possuíam governança corporativa. Adicionalmente, das 1169 variáveis que o modelo previu possuir governança corporativa, apenas 28 não possuíam (Tabela5). Esses resultados levam a argumentação de que através de 6 indicadores, dos quais 3 são financeiros e os outros 3 são econômicos, é possível prever se a empresa possui ou não governança corporativa.

## 5. Considerações Finais

Os resultados obtidos por meio da regressão logística apresentaram boa qualidade de ajuste, houve maior índice de acertos na predição das empresas que possuem governança corporativa, o que indica bom ajuste do modelo. O modelo geral teve aproximadamente 81% de acerto em sua estimação sobre as empresas que fazem parte dos níveis diferenciados de governança corporativa da Bovespa, e conseqüentemente possuem governança corporativa. Sendo que o nível de previsão do modelo para as empresas que possuem governança corporativa foi de 97,6%, enquanto o nível de previsão para aquelas que não possuem governança corporativa foi de 18,20%.

Diante do exposto, o modelo logístico mostra-se como um instrumento promissor para governança corporativa, podendo ser aprimorado com a inclusão de outras variáveis e podendo além disto construir outros modelos que demonstrem a previsão para os níveis diferenciados de governança corporativa existentes. Por fim, evidencia-se as limitações do trabalho que se referem ao caráter probabilístico do método empregado e ao horizonte temporal de análise (curto prazo). Sugere-se para pesquisas futuras o emprego de outras metodologias, tais como redes neurais, análise discriminante, ARCH, entre outros, para a determinação da adoção de práticas de governança corporativa. Além disso, contemplar as diferentes tipologias apresentadas nos níveis diferenciados de governança corporativa.

## Referências

- [1] ALMEIDA, M. A.; SANTOS, J. F.; FERREIRA, L. F. F. M.; TORRES, F. J. F. *Evolução da qualidade das práticas de governança corporativa: um estudo das empresas Brasileiras de capital aberto não listadas em bolsa*. Rev. adm. contemp. vol.14 no.5 Curitiba Sept./Oct. 2010
- [2] BOLSA DE VALORES DE SÃO PAULO (BOVESPA). *Empresas listadas*. Disponível em: <<http://bovespa.com.br/>>. Acesso em: 14 dez. 2013.
- [3] IQUIAPAZZA, R. A.; GRUBERGER, D. J.; SANTOS, I. T. *A Endogeneidade da estrutura de propriedade, governança corporativa e criação de valor na BOVESPA no período 2002-2005*. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 31, 2007, Rio de Janeiro. Anais... ENANPAD, XXXI. Rio de Janeiro: ANPAD, 2007.

# Sequential Monte Carlo methods with move and reweighting schemes

Reinaldo Marques and Geir Storvik

**Abstract** Sequential Monte Carlo methods are widely used to deal with the intractability of complex models including state space models. Their aim is to approximate the distribution of interest by a set of properly weighted samples. To control the weight degeneracy, the resample step has been proposed as an inexpensive alternative to avoid the collapse of particle filter algorithms. When the sample becomes too poor with successive use of resample steps, MCMC moves have been added in particle filter algorithms in order to make the identical samples diverge. In this work we consider strategies where we first perform a moves step, and then we update the weights for reweighting the particles. The validity of this approach is based on the commonly used trick of working on an artificial extended distribution having the target distribution as marginal combined with the use of backwards kernels. By updating the weights via a diversification step, this approach can make their empirical distribution less skewed increasing the effective sample size.

## 1 Introduction

State space models provide flexible representations for stochastic dynamical systems in which they can encapsulate many real problems (in time or space-time). This class of models is expressed in terms of an initial distribution,  $\mathbf{x}_1 \sim \pi_\theta(\mathbf{x}_1)$ , a state Markov model,  $\mathbf{x}_t | \mathbf{x}_{1:t-1} \sim \pi_\theta(\mathbf{x}_t | \mathbf{x}_{t-1})$ , and the observation model,  $\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{x}_{1:t} \sim \pi_\theta(\mathbf{y}_t | \mathbf{x}_t)$ . All static parameters are represented by  $\theta \in \Theta$ , and for simplicity of notation we assume each  $\mathbf{x}_t$  to have a common sample space  $\mathcal{X} \subset \mathbf{R}^{p_x}$ .

Given the data  $\mathbf{y}_{1:t}$  up to time  $t$  and assuming  $\theta$  is known, inference focuses on the posterior distribution

---

Reinaldo Marques  
University of Oslo and Statistics for Innovation Center, Norway, e-mail: ramarque@math.uio.no

Geir Storvik  
University of Oslo and Statistics for Innovation Center, Norway e-mail: geirs@math.uio.no

$$\pi_t(\mathbf{x}_{1:t}) \equiv \pi_\theta(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) \propto \pi_\theta(\mathbf{x}_1)\pi_\theta(\mathbf{y}_1|\mathbf{x}_1) \prod_{k=2}^t \pi_\theta(\mathbf{x}_k|\mathbf{x}_{k-1})\pi_\theta(\mathbf{y}_k|\mathbf{x}_k). \quad (1)$$

When  $\pi_t(\mathbf{x}_{1:t})$  is intractable, sequential Monte Carlo methods can be applied to carry out an approximate inference. Their respective algorithms –called *particle filters* (PF)– are exhaustively used to deal with the computational intractability of general state space models (see [7], [1], [2] and the references therein). The main idea is approximated the posterior by a set of properly weighted samples. The population of samples are called particles, and they are typically generated sequentially from some low-dimensional conditional distributions

$$q_t(\mathbf{x}_{1:t}) = q(\mathbf{x}_1) \prod_{k=2}^t q(\mathbf{x}_k|\mathbf{x}_{k-1}). \quad (2)$$

Assuming that  $q_t(\mathbf{x}_{1:t}) > 0$  for all  $\mathbf{x}_{1:t}$  with  $\pi_t(\mathbf{x}_{1:t}) > 0$ , then based on sequential importance sampling ideas, particle weights are defined as

$$w_t(\mathbf{x}_{1:t}) \equiv \frac{\pi_t(\mathbf{x}_{1:t})}{q_t(\mathbf{x}_{1:t})} \propto w_{t-1} \frac{\pi_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})\pi_\theta(\mathbf{y}_t|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \quad (3)$$

allowing for recursive computation. Estimation is usually based on normalized weights, making the proportionality constant unnecessary to compute.

A well-known problem in SMC methods is *weight degeneracy*, also called sample impoverishment. This problem is related to the increasing variance of the particle weights over time. In order to attenuate the degeneracy problem, [5] suggested to add resampling steps. Later, [4] proposed to add MCMC moves after the resampling step to reduce the sample impoverishment. This approach is called the resample-move (RM) algorithm, and the main idea is to create a greater diversity in the sample by rejuvenating the particles via a combination of sequential importance resampling and MCMC sampling steps. For this scheme, a Markov kernel  $K_t(\cdot|\mathbf{x}_{1:t})$  with  $\pi_t(\mathbf{x}_{1:t})$  as the stationary distribution is designed to draw samples after the resample steps.

## 2 Particle move-reweighting strategies

Assume  $\mathbf{x}_{1:t} \sim q_t(\mathbf{x}_{1:t})$  is followed by a move  $\mathbf{x}_{1:t}^* \sim K_t(\mathbf{x}_{1:t}^*|\mathbf{x}_{1:t})$  where  $K_t$  is  $\pi_t$ -invariant. As suggested by [3] and [8] define an extended target distribution as  $\bar{\pi}_t(\mathbf{x}_{1:t}^*, \mathbf{x}_{1:t}) \equiv \pi_t(\mathbf{x}_{1:t}^*)h_t(\mathbf{x}_{1:t}|\mathbf{x}_{1:t}^*)$ , where  $h_t$  is an artificial density/ backward kernel that integrates to one. Following ordinary importance sample theory working on the enlarged space, we obtain the following result:

**Proposition 1.** *Let  $\{(\mathbf{x}_{1:t}^i, w_t(\mathbf{x}_{1:t}^i)), i = 1, \dots, N\}$  be a properly weighted sample with respect to  $\pi_t$  where the particles are generated from  $q_t$ . Assume we make a move by a  $\pi_t$ -invariant MCMC kernel  $K_t$  to  $\mathbf{x}_{1:t}^*$  and update the weights by*

$$\bar{w}_t(\mathbf{x}_{1:t}; \mathbf{x}_{1:t}^*) = w_t(\mathbf{x}_{1:t}) \times r_t(\mathbf{x}_{1:t}; \mathbf{x}_{1:t}^*)$$

where

$$r_t(\mathbf{x}_{1:t}; \mathbf{x}_{1:t}^*) = \frac{\pi_t(\mathbf{x}_{1:t}^*) h_t(\mathbf{x}_{1:t} | \mathbf{x}_{1:t}^*)}{\pi_t(\mathbf{x}_{1:t}) K_t(\mathbf{x}_{1:t}^* | \mathbf{x}_{1:t})},$$

and  $h_t$  is a density such that  $\{(\mathbf{x}_{1:t}^*, \mathbf{x}_{1:t}) : \pi_t(\mathbf{x}_{1:t}^*) h_t(\mathbf{x}_{1:t} | \mathbf{x}_{1:t}^*) > 0\}$  is a subset of  $\{(\mathbf{x}_{1:t}^*, \mathbf{x}_{1:t}) : q_t(\mathbf{x}_{1:t}^*) K_t(\mathbf{x}_{1:t} | \mathbf{x}_{1:t}^*) > 0\}$ . Then

$$\{(\mathbf{x}_{1:t}^{*i}, \bar{w}_t(\mathbf{x}_{1:t}^i; \mathbf{x}_{1:t}^{*i})), i = 1, \dots, N\}$$

becomes proper with respect to  $\pi_t$ .

For a proof, see [6].

By adding the MCMC move and updating the particle weights, we have that, for any density  $h$ ,  $\bar{w}_t$  is a proper weight for  $\mathbf{x}_{1:t}^*$  with respect to  $\pi_t$ . Even though we are working with an extended space, the MCMC move does not modify the unbiased property. The main point of this sampling scheme is to highlight the multiple choices of proper weight functions  $\bar{w}_t(\mathbf{x}_{1:t}; \mathbf{x}_{1:t}^*)$  for any given kernel  $K_t$ . For specific choice of  $h_t$ ,  $\bar{w}_t$  reduces to  $w_t$ . Also, for clever choices of the artificial density, we can obtain that  $Var[\bar{w}_t] < Var[w_t]$  allowing us to reduce the weight degeneracy and simultaneously to create sample diversity. Some special choices of the artificial density and moves from arbitrary transitions kernels are discussed in [6].

When we have a properly weighted sample, there are some interesting SMC schemes to design, taking into different types of moves. In short, we can adduce at least four strategies for (re)-weighting the particles in an SMC framework:

- s.1 Start with an equally weighted sample, move the particles via an MCMC kernel and keep the equal particle weights (the *resample-move* approach);
- s.2 Start with a properly weighted sample, move the particles via an MCMC kernel and keep the same particle weights;
- s.3 Start with a properly weighted sample, move the particles via an MCMC kernel and update the particle weights; or
- s.4 Start with a properly weighted sample, move the particles via an arbitrary kernel and update the particle weights.

**Table 1** Results of resample-move (RM) and move-reweighting (MR) algorithms. The table shows the average over time of the effective sample size (ESS) and root mean square root (RMSE) for 50 particles and sample size equal to 200.

Algorithms	ESS	RMSE
RM	26%	94.54
MR	67%	84.40
MR-RM	70%	83.38

The Table 1 summarizes results for the following state space model:  $y_t \sim \text{Poisson}(e^{\mathbf{b}\mathbf{x}_t + \lambda})$  and  $\mathbf{x}_t \sim \mathcal{N}(\mathbf{F}\mathbf{x}_{t-1}, \Sigma)$ . All static parameter we assume to be known. The Table shows that updating the weights via a MCMC move, we can obtain a considerable increase in the effective sample size and an improvement in the predictive performance.

### 3 Closing remarks

In this work, we propose a flexible strategy that allows for MCMC moves without the need of a preliminary resampling step. Following the MCMC moves, we update the particle weight taking into account the diversification step and that MCMC moves give particles closer to the target distribution. Our approach allows a great flexibility to reweight the particles using a proper weight with respect to the right distribution. The main effect of this is that we can increase the effective sample size, consequently the resample stages can be delayed.

**Acknowledgements** We gratefully acknowledge financial support from CAPES-Brazil and Statistics for Innovation Center, in Norway.

### References

1. C. Andrieu, A. Doucet and R. Holenstein. **Particle Markov chain Monte Carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*; 2010; 72(3); pp. 269-342.
2. N. Chopin, P. Jacob and O. Papaspiliopoulos. **SMC<sup>2</sup>: an efficient algorithm for sequential analysis of state space models**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*; 2012; doi: 10.1111/j.1467-9868.2012.01046.x
3. P. Del Moral, A. Doucet, and A. Jasra. **Sequential Monte Carlo samplers**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*; 2006; 68(3); pp. 411-436.
4. W. Gilks and C. Berzuini. **Following a moving target - Monte Carlo Inference for dynamic Bayesian models**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*; 2001; 63(1); pp. 127-146.
5. N. Gordon and D. Salmond. **Novel approach to nonlinear/non-Gaussian Bayesian state estimation**. *Radar and Signal Processing, IEE Proceedings F*; 1993; 140(2); pp. 107-113.
6. R. Marques and G. Storvik. **Online inference for dynamical models: the move-reweighting particle filter** (*In preparation*).
7. C. Robert and G. Casella. **Monte Carlo statistical methods**. Springer Verlag; 2004.
8. G. Storvik. **On the Flexibility of Metropolis–Hastings Acceptance Probabilities in Auxiliary Variable Proposal Generation**. *Scandinavian Journal of Statistics*; 2011; 38(2); pp. 342-358.

# A importância dos experimentos em faixas na Sensometria: o caso do queijo Minas Padrão com inulina

Laís Brambilla Storti<sup>1 4</sup>

Eric Batista Ferreira<sup>2 4</sup>

Celeide Pereira<sup>3 4</sup>

**Resumo:** Alimentos funcionais são aqueles desenvolvidos para trazer benefícios à saúde, além de possuírem os nutrientes já tradicionais. A inulina é um importante prebiótico que pode tornar o queijo Minas Padrão um alimento funcional. Por outro lado, a adição de inulina não deve afetar a qualidade sensorial do queijo. Um experimento em faixas foi desenvolvido para verificar o impacto sensorial (sabor, aparência, textura e cor) da adição de 0%, 2% e 4% de inulina ao queijo Minas Padrão, que foi avaliado por provadores treinados aos 0, 15, 30 e 45 dias de maturação. A função `faixas()` foi programada no pacote `ExpDes.pt` do software R com a finalidade de facilitar a análise desse tipo de experimento. O uso da dose 2% de inulina é aconselhada do ponto de vista sensorial, por não afetar a qualidade do queijo durante o período de maturação. De forma geral, os experimentos em faixa são muito importantes em Sensometria, devido a enorme importância de provadores treinados acompanharem a evolução temporal de variáveis sensoriais.

**Palavras-chave:** *Experimentos em faixas, Análise sensorial, Queijo Minas Padrão.*

## 1 Introdução

Por volta da década de 1980 o conceito de alimento funcional foi proposto no Japão fazendo referência àqueles alimentos que ao serem usados como parte da dieta normal trazem benefícios a saúde (COSTA; ROSA, 2010). Porém, nas últimas décadas, este conceito voltou-se para o uso de aditivos alimentares que podem influenciar na composição da microbiota intestinal. Esses aditivos são probióticos, que contêm bactérias que propiciam certo benefício ao hospedeiro e que agem no equilíbrio da microbiota intestinal, e prebióticos que não são digeríveis e também trazem benefícios ao hospedeiro por estimularem a proliferação de bactérias (probióticos) no intestino (SAAD, 2006).

---

<sup>1</sup>Universidade Federal de Alfenas (Unifal-MG). Email: [lahstorti@gmail.com](mailto:lahstorti@gmail.com)

<sup>2</sup>Professor do Instituto de Ciências Exatas (ICEX) - Universidade Federal de Alfenas (Unifal-MG). Email: [eric.ferreira@unifal-mg.edu.br](mailto:eric.ferreira@unifal-mg.edu.br)

<sup>3</sup>Professora da Universidade Tecnológica Federal do Paraná (UTFPR), Campus de Medianeira, PR. Email: [celeide@utfpr.edu.br](mailto:celeide@utfpr.edu.br)

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

Um exemplo de prebiótico utilizado para este fim é a inulina, que é um carboidrato formado por uma cadeia de moléculas de frutose e uma molécula de glicose terminal (SILVA, 1996). Um exemplo da utilização de inulina é no processo de fabricação do queijo Minas Padrão.

Com a adição de inulina é esperado que características funcionais importantes sejam trazidas ao queijo Minas Padrão, entretanto, é fundamental que tal adição não prejudique as características sensoriais do queijo. Por isso, os experimentos sensoriais são recomendados na fase de desenvolvimento do produto, principalmente utilizando-se provadores treinados que acompanham a evolução do queijo ao longo de sua vida de prateleira. É justamente nesse contexto que são indicados os *experimento em faixas*.

O experimento em faixas, ou *split-blocks*, permite realizar um teste simultâneo de mais de um fator, onde existe uma restrição de aleatorização (como entre os níveis de tempo ou espaço). O fator principal é colocado nas parcelas e o secundário nas subparcelas. Neste tipo de experimento existem três tipos de erro: um associado as parcelas (erro a), o outro às subparcelas (erro b) e o último associado à interação da parcela com a subparcela (erro c) (DIAS; BARROS, 2009).

Os experimentos sensoriais podem ser analisados em faixas, pois os provadores são blocos típicos e o tempo não é passível de aleatorização, logo são dispostos em faixas. Assim, basta mais um fator de interesse para utilizar o experimento em faixas.

No contexto sensorial, o fator mais interessante de se alocar às subparcelas é o tempo, especificamente, a vida de prateleira. Como o outro fator é alocado à parcela, a interação bem estimada permitirá verificar se os níveis do tratamento adotado proporcionarão resultados homogêneos ao longo da vida do produto, do ponto de vista sensorial.

## 2 Objetivo

Este trabalho tem como objetivos mostrar a importância da utilização de experimentos em faixas na análise sensorial e inferir sobre a influência da inulina nas características sensoriais do queijo Minas Padrão ao longo da sua vida de prateleira.

## 3 Metodologia

De acordo com Pereira (2005), o experimento foi conduzido no Laboratório de Laticínios do Departamento de Ciência dos alimentos (DCA/UFLA) e o queijo foi fabricado no Laticínio Verde Campo, em Lavras, MG. Foram adicionadas ao queijo as doses de 0% (testemunha), 2% e 4% de inulina e as amostras de queijo foram coletadas nos tempos 0, 15, 30, 45 dias de maturação e submetidas às análises sensoriais, onde os 15 provadores treinados avaliaram as amostras para os atributos sabor, textura, cor e aparência durante dois horários, de manhã e a tarde (PEREIRA,

2005).

Já para a análise estatística dos dados foi utilizado um delineamento em faixas e foi feito o uso da função `faixas()` do pacote `ExpDes.pt` (FERREIRA; CAVALCANTI; NOGUEIRA, 2013) do software R (R CORE TEAM, 2014). A função `faixas` realiza a análise de variância e caso a interação dos fatores for significativa realiza o desdobramento da interação, onde são realizados o teste de Tukey para médias e a análise de regressão.

O modelo estatístico para experimentos em faixas é dado por:  $Y_{ijk} = \mu + \alpha_i + b_j + \varepsilon_{ij} + \beta_k + \varepsilon_{jk} + \alpha\beta_{ik} + \varepsilon_{ijk}$ , no qual  $\mu$  é a média geral;  $\alpha_i$  é o efeito do tratamento na parcela, no caso do queijo minas padrão é a inulina;  $b_j$  é o efeito de blocos, neste caso os provadores;  $\varepsilon_{ij}$  é o erro aleatório do fator da parcela;  $\beta_k$  é o efeito nas subparcelas, no experimento em questão é o tempo;  $\varepsilon_{jk}$  é o erro aleatório do fator da subparcela;  $\alpha\beta_{ik}$  é o efeito da interação entre os fatores da parcela e da subparcela (inulina  $\times$  tempo);  $\varepsilon_{ijk}$  é o erro aleatório da interação.

As esperanças de quadrados médios (componentes de variância) para o experimento em faixas, adaptados de Federer e King (2007), são expressadas na Tabela 1.

Tabela 1: Componentes de variância do experimento em faixas

Fonte de variação	$E[QM]$
$\alpha_i$	$JK\phi_\alpha + K\sigma_{ij}^2 + \sigma_{ijk}^2$
$b_j$	$IK\sigma_\beta + K\sigma_{ij}^2 + I\sigma_{jk}^2 + \sigma_{ijk}^2$
$\varepsilon_{ij}$	$K\sigma_{ij}^2 + \sigma_{ijk}^2$
$\beta_k$	$IJ\phi_\beta + I\sigma_{jk}^2 + \sigma_{ijk}^2$
$\varepsilon_{jk}$	$I\sigma_{jk}^2 + \sigma_{ijk}^2$
$\alpha\beta_{ik}$	$J\phi_{\alpha\beta} + \sigma_{ijk}^2$
$\varepsilon_{ijk}$	$\sigma_{ijk}^2$

A partir dos componentes de variância pode-se determinar o denominador do teste F. Neste caso, observa-se que a fonte de variação  $\alpha_i$  é testada com o erro  $\varepsilon_{ij}$ , o efeito  $\beta_k$  com o erro  $\varepsilon_{jk}$  e a interação  $\alpha\beta_{ik}$  é testada com o erro  $\varepsilon_{ijk}$ . No caso do efeito de bloco ( $b_j$ ), ele é testado com um erro combinado  $QM_{comb} = QM_\alpha + QM_\beta - QM_{\alpha\beta}$ .

## 4 Resultados e Discussão

Inicialmente foi realizada a análise de variância utilizando a função `faixas()`, onde foi possível observar que para os atributos aparência e textura a interação *tempo*  $\times$  *inulina* foram significativas a 5% de significância. Já para os atributos cor e sabor a interação foi não significativa. Além disso, foi possível observar que, para o atributo sabor, o efeito de bloco foi significativo a 5% de significância, o que representa que os provadores não concordam entre si (Figura 1).

Através da análise das médias dos atributos sensoriais ao longo dos tempos de armazenamento e doses de inulina estudados pode-se observar que na maioria dos atributos as médias

```

-----
$`Quadro da análise de variância
-----

```

	GL	SQ	QM	Fc	Pr(>Fc)
Bloco	4	88.475	22.119	6.3537	0.01328 *
Inulina	2	5.308	2.654	1.3354	0.31591
Erro a	8	15.900	1.988		
Tempo	3	17.646	5.882	2.7018	0.09241 .
Erro b	12	26.125	2.177		
Inulina*Tempo	6	3.392	0.565	0.8272	0.56034
Erro c	24	16.400	0.683		
Total	59	173.246	36.068		

```

-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----
CV 1 = 21.99928 %
CV 2 = 23.02462 %
CV 3 = 12.89945 %

```

Figura 1: Quadro da análise de variância do atributo sabor

são consideradas estatisticamente iguais. Porém, para o atributo textura a partir do 15º dia de armazenamento há diferença na textura com relação as diferentes doses de inulina e para o atributo aparência nota-se essa diferença a partir do 30º dia de armazenamento.

Para ilustrar a diferença existente entre tempo nos atributos aparência e textura, realizou-se a análise de regressão. Na Figura 2(a) pode-se observar que na dose 2% de inulina o atributo aparência tem um comportamento decrescente com o passar do tempo (linha tracejada). Já para a dose 4% de inulina (linha cheia) a aparência tem comportamento decrescente até 20º dia, quando começa crescer novamente.

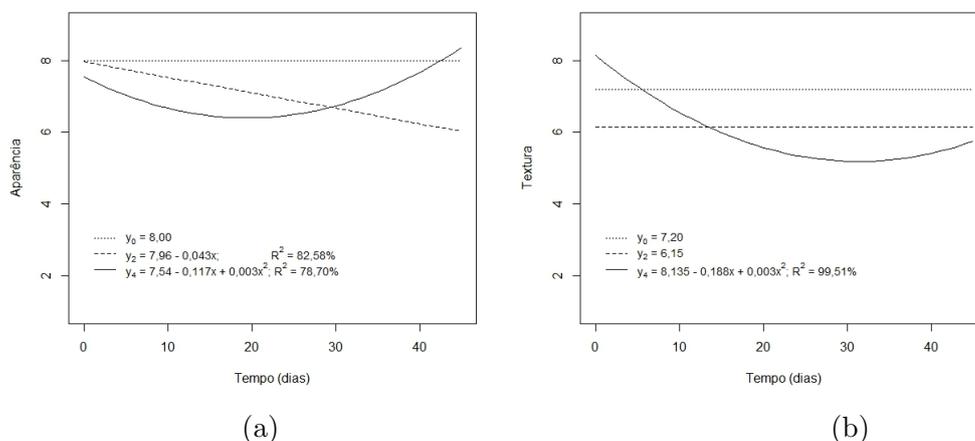


Figura 2: Gráficos dos modelos modelos de regressão ajustados para os atributos Aparência (a) e Textura (b), para as doses 0%, 2% e 4% de inulina.

Com relação ao atributo textura, na Figura 2(b) é possível observar que com as doses 0% e 2% de inulina (linhas pontilhada e tracejada, respectivamente) a textura do queijo ficou constante. Já para a dose de 4% (linha cheia) pode-se notar um comportamento decrescente até o 30º dia.

Pelo fato do queijo Minas Padrão ser consumido a partir do 10º dia, apesar de sua maturação acontecer até o 30º dia, é possível concluir que a dose 2% de inulina é a mais indicada, uma vez que preserva as qualidades sensoriais do queijo. A dose 4% parece interferir na qualidade

sensorial a partir de 20 dias de maturação e a dose 0%, embora não interfira na qualidade sensorial, não torna o alimento funcional.

## 5 Conclusões

Podemos observar que, através das análises realizadas, a dose 2% de inulina é a mais indicada, pois propicia o aspecto funcional preservando as características sensoriais ao longo dos tempos analisados.

A utilização do experimento em faixas foi importante por permitir a realização da análise dos dados de uma só vez, o que torna a análise mais prática, rápida e de fácil compreensão e utilização, principalmente para usuários não estatísticos.

Além disso, na análise sensorial, é muito comum o interesse em estudar o tempo de prateleira dos alimentos. Logo, é possível utilizar o experimento em faixas, uma vez que os blocos serão sempre os provadores e o tempo, como não é passível de aleatorização, será sempre alocado na subparcela, propiciando uma análise mais rápida e direta dos dados.

## Referências

- [1] COSTA, N. M. B.; ROSA, C. O. B. *Alimentos funcionais: componentes bioativos e efeitos fisiológicos*. Rio de Janeiro: Editora Rubio, 2010.
- [2] DIAS, L. A. S.; BARROS, W. S. *Biometria experimental*. Viçosa: UFV, 2009.
- [3] FERREIRA, E. B.; CAVALCANTI, P. P.; NOGUEIRA, D. A. *ExpDes.pt: Pacote Experimental Designs (Portuguese)*. R package version 1.1.2. 2013.
- [4] FEDERER, W. T.; KING, F. *Variations on split plot and split block experiment designs*. New Jersey: Wiley, 2007.
- [5] PEREIRA, C. *Elaboração de queijo minas padrão com adição de probióticos e prebióticos*. 2005. 123f. Dissertação (Mestrado em Ciências dos Alimentos), Universidade Federal de Lavras, Lavras, 2005.
- [6] R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. 2014.
- [7] SAAD, S. M. I. Probióticos e prebióticos: o estado da arte *Revista Brasileira de Ciências Farmacêuticas*, vol. 42, n. 1, jan./mar., 2006.
- [8] SILVA, R. F. Use of inulin as a natural texture modifier. *Cereal Foods World*, St Paul, v. 41, n. 10, p. 792-795, 1996.

# Otimização com respostas simultâneas na selagem da noz macadâmia

**Bruna de Oliveira Gonçalves**<sup>1 5</sup>

**Lislaine Cristina Cardoso**<sup>2 5</sup>

**Elcio do Nascimento Chagas**<sup>3 5</sup>

**Eric Batista Ferreira**<sup>4 5</sup>

A abordagem comumente utilizada na metodologia de superfície de resposta consiste na análise de uma única variável de cada vez, mantendo-se constantes as restantes. Entretanto, quando um experimento é executado, torna-se necessário medir várias respostas em relação aos fatores. Determinar um conjunto de níveis dos fatores ótimos para as respostas não é um problema simples, pois dificilmente os níveis dos fatores que são ótimos para uma determinada resposta, serão ótimos para as outras. Por este motivo é importante considerar, de forma conjunta, o efeito dos fatores. A metodologia denominada otimização simultânea de várias respostas é utilizada em experimentos que forneçam duas ou mais respostas de interesse. Dessa forma, torna-se desejável obter uma resposta ótima, que satisfaça todas as variáveis dependentes envolvidas. Os resultados da aplicação no estudo das variáveis que afetam a selagem da noz macadâmia nas características do filme de cobertura formado sobre a noz mostram uma solução global, que representa a melhor combinação, no sentido de proporcionar a resposta mínima de forma simultânea para as variáveis analisadas.

**Palavras-chave:** *Função desirability, Metodologia de superfície de resposta.*

## 1 Introdução

A metodologia de superfície de resposta (MSR) consiste em uma coleção de técnicas estatísticas e matemáticas útil para desenvolvimento, melhora e otimização de processos (MYERS; MONTGOMERY; ANDERSON-COOK, 2009).

A abordagem comumente utilizada na MSR, consiste na análise de uma única variável de cada vez, mantendo-se constantes as restantes. Entretanto, quando um experimento é executado, torna-se necessário medir várias respostas em relação aos fatores. Determinar um conjunto de

<sup>1</sup>ICEX - UNIFAL. Universidade Federal de Alfenas, Alfenas-MG. E-mail: *brunag-25@hotmail.com*

<sup>2</sup>ICEX - UNIFAL. Universidade Federal de Alfenas, Alfenas-MG. E-mail: *lcc-mat.uems@hotmail.com*

<sup>3</sup>IFES- Instituto Federal do Espírito Santo, *Campus* de Alegre -ES. E-mail: *enchagas@ifes.edu.br*

<sup>4</sup>ICEX - UNIFAL. Universidade Federal de Alfenas, Alfenas-MG. E-mail: *eric.ferreira@unifal-mg.edu.br*

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

níveis dos fatores ótimos para as respostas não é um problema simples, pois dificilmente os níveis dos fatores que são ótimos para uma determinada resposta, serão ótimos para as outras. Por este motivo é importante considerar, de forma conjunta, o efeito dos fatores.

Derringer e Suich (1980) criaram a metodologia denominada otimização simultânea de várias respostas, utilizada em experimentos que forneçam mais de uma resposta de interesse. Os autores propuseram uma nova medida denominada função *desirability* (função desejabilidade) baseada na ideia de que a qualidade de um produto ou processo que tem múltiplas respostas, depende de encontrar as variações de processo que atendam a todas as restrições impostas. O método verifica, sem comprometer ou desprezar as condições operacionais nas quais o experimento foi realizado, as condições que fornecem o valor da resposta mais desejável.

## 2 Objetivos

Este trabalho tem como objetivo, aplicar a metodologia de otimização com respostas simultâneas no estudo das variáveis que afetam a selagem da noz macadâmia nas características do filme de cobertura formado sobre a noz.

## 3 Metodologia

A análise de problemas com respostas múltiplas tem sido alvo de um interesse crescente em diferentes áreas do conhecimento. A resolução deste tipo de problemas exige que se modele cada uma das respostas que se pretende otimizar por uma função que descreva a denominada superfície de resposta permitindo estimar o valor da resposta dentro do intervalo de variação definido para as variáveis envolvidas no estudo. Consiste em transformar as equações que modelam cada respostas em funções individuais, e depois proceder à otimização de uma função global (Global Desirability, D) que é descrita em termos das funções individuais. A otimização simultânea de várias respostas transforma-se assim na otimização de uma única função.

A otimização das respostas envolvidas no estudo é efetuada por meio da maximização da função *desirability* global  $D$  de acordo com a expressão (DERRINGER; SUICH, 1980)

$$D = \left( \prod_{j=1}^k d_j \right)^{1/k} \quad (1)$$

em que  $k$  corresponde ao número de respostas,  $0 \leq d_j \leq 1$  e, por conseguinte,  $0 \leq D \leq 1$ .

A interpretação da função desejável  $d_j$  ( $j = 1, \dots, k$ ) para cada variável resposta, considera uma escala de 0 à 1. Para esses valores respectivamente, interpreta-se que o ponto ótimo

pesquisado seja completamente indesejável, face aos objetivos propostos, ao passo que, no maior valor assumido ( $d = 1$ ) o resultado é satisfatório.

Quando o valor-alvo ( $Y_T$ ) de uma resposta predita ( $\hat{Y}$ ) está entre um valor máximo ( $Y_U$ ) e um valor mínimo ( $Y_L$ ), a correspondente *desirability* individual, ( $d$ ), é definida por

$$d = \begin{cases} 0, & \text{para } \hat{Y} < Y_L \\ \left( \frac{\hat{Y} - Y_L}{Y_T - Y_L} \right)^s, & \text{para } Y_L < \hat{Y} < Y_T \\ \left( \frac{\hat{Y} - Y_U}{Y_T - Y_U} \right)^t, & \text{para } Y_T < \hat{Y} < Y_U \\ 0, & \text{para } \hat{Y} > Y_U \end{cases} \quad (2)$$

em que:  $Y_L$  é o menor valor possível para qualquer função desejável;  $Y_U$  é o maior valor possível para qualquer função desejável e  $Y_T$  é o valor específico (valor-alvo), para a resposta de interesse  $Y_i$ . Com  $Y_L < Y_U < Y_T$ .

Neste estudo a resposta de interesse é mínima. Assim utiliza-se a expressão

$$d = \begin{cases} 1, & \text{para } \hat{Y} < Y_T \\ \left( \frac{\hat{Y} - Y_U}{Y_T - Y_U} \right)^s, & \text{para } Y_T < \hat{Y} < Y_U \\ 0, & \text{para } \hat{Y} > Y_U \end{cases} \quad (3)$$

em que  $Y_T$  é o valor mínimo para a resposta. (ENGINEERING STATISTICS HANDBOOK, 2014).

### 3.1 Exemplo utilizado para aplicação do método - avaliação das variáveis que afetam a selagem da noz macadâmia nas características do filme de cobertura formado sobre a noz macadâmia.

Neste experimento empregou-se um delineamento Composto Central Rotacional de segunda ordem, com base na MSR, a fim de se avaliar a influência da concentração de selante (33,2 a 68%), rotação da drageadeira (23,2 a 56,8 rpm) e número de cargas de solução selante aplicadas (1 a 11) (variáveis de controle) nas características do filme de cobertura formado sobre a macadâmia. Maiores detalhes podem ser encontrados em (FADINI et. al.,2006). Para a análise dos resultados, utilizou-se o pacote *desirability* do software R (R CORE TEAM, 2014).

## 4 Resultados e Discussão

Analisando os sinais dos autovalores da superfície de resposta ajustada para a porcentagem de área descoberta, tem-se que o ponto estacionário representa um ponto de sela dentro da superfície de resposta (nem máximo, nem mínimo). Na superfície de resposta apresentada na Figura 1 pode-se notar que para os fatores rotação de drageadeira versus solução de goma acácia a superfície apresentou um ponto de sela. Portanto, tornou-se necessário a utilização do método *Ridge* para inferir sobre a determinação do mínimo local para esta resposta dentro da região de experimentação estudada (MYERS; MONTGOMERY; ANDERSON-COOK, 2009).

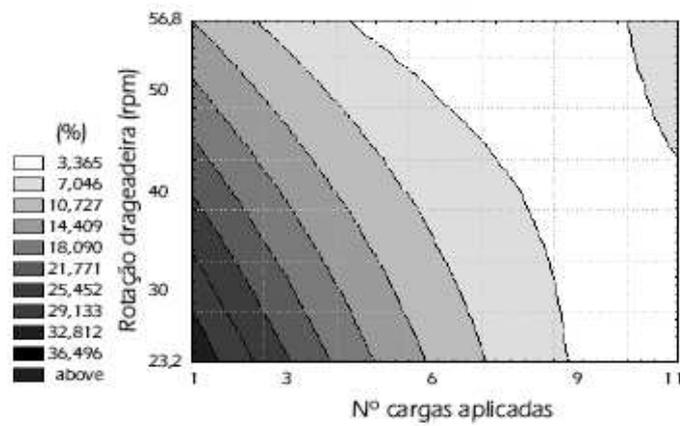


Figura 1: Superfície de resposta para a porcentagem de área descoberta.

As estimativas dos parâmetros do modelo ajustado por meio do método *Ridge* encontram-se apresentadas pela equação (4)

$$\hat{Y} = 10,69 + 2,60x_1 - 0,95x_2 - 2,06x_3 + 1,45x_1^2 + 0,42x_3^2 + 0,45x_2x_3 \quad (4)$$

Da mesma forma, os resultados também mostraram que o ponto estacionário representa um ponto de sela para a espessura do filme de selagem. Assim as estimativas dos parâmetros do modelo ajustado por meio do método *Ridge* encontram-se apresentadas pela equação (5)

$$\hat{Y} = 73,83 + 2,94x_1 - 1,96x_2 + 12,16x_3 + 1,45x_1^2 + 2,26x_2^2 - 3,24x_3^2 + 0,70x_1x_2 - 0,08x_1x_3 \quad (5)$$

Tendo em vista que os resultados encontrados para o ponto estacionário resultaram num ponto de sela, nota-se que a análise estatística feita de forma univariada irá requerer um esforço maior para validação dos resultados. Assim, pode-se justificar a análise das respostas de forma simultânea de forma que satisfaça os objetivos propostos para a produção de uma noz macadâmia

com baixa porcentagem de área descoberta (aproximadamente 5%) como proposto por Fadini et al. (2006).

A solução global, descrita que representa a melhor combinação de forma a proporcionar uma resposta mínima de forma simultânea para as variáveis respostas foi rotação de 42 rpm, 5 aplicações de solução selante com 41% de goma acácia.

Os autores concluíram que a melhor condição parece ser aquela com uma rotação de 54 rpm, 5 aplicações de solução selante com 43% de goma acácia. No que diz respeito à espessura do filme de cobertura foi observado que há uma pequena tendência de diminuição da espessura com o aumento da rotação, para um mesmo número de cargas e que quanto maior a concentração de solução selante, maior a espessura observada. No entanto, isto implica na obtenção de macadâmias com maior porcentagem de área descoberta, o que poderá ser prejudicial para a estabilidade do produto.

## 5 Conclusões

Os resultados mostram uma solução global, que representa a melhor combinação, no sentido de proporcionar a resposta mínima de forma simultânea para as variáveis analisadas. A praticidade da realização desse método, mediante a um experimento que proporcione várias respostas, de simplifica toda a análise estatística, permitindo-se determinar um único ponto comum a todas as variáveis, respeitando os diferentes objetivos propostos.

## Referências

- [1] DERRINGER, G.; SUICH, R. Simultaneous Optimization of Several Response Variables. *Journal of Quality Technology*, 12(4), p. 214-219, 1980.
- [2] ENGINEERING STATISTICS HANDBOOK. *Multiple responses: The desirability approach*. Disponível em: <<http://www.itl.nist.gov/div898/handbook/pri/section5/pri5322.htm>>. Acesso em: 30 de julho de 2014.
- [3] MYERS, R. H.; MONTGOMERY, D. C.; ANDERSON-COOK, C. M. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 3. ed. Wiley, 2009
- [4] R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. Disponível em: <http://www.R-project.org/>. Acesso em 30 julho. 2014.

# Estudo de quantidades de aves, bovinos e suíno abatidos durante os anos de 1975 a 2013 usando séries temporais

Valiana Alves Teodoro <sup>1 4</sup>

Djair Durand Ramalho Frade <sup>2 4</sup>

Taciana Villela Savian <sup>3 4</sup>

**Resumo:** Como a produção de carne bovina, suína e de aves representam um importante segmento do agronegócio brasileiro, é de suma importância realizar estudos referentes a estas atividades que muito tem contribuído para o crescimento do país. Com este trabalho pretende-se utilizar modelos de séries temporal para o estudo das séries de quantidade de aves, bovinos e suínos. As séries originais de quantidades de aves, bovinos e suínos abatidos tem uma tendência de crescimento e um acréscimo na variância a medida que o tempo passa. Com isso, uma transformação logarítmica foi adequada para estabilizar esta variância, assim, passa-se a analisar o logaritmo de cada série, com isso passa a se estudar a série transformada. Estudando o comportamento do gráfico de ACF das séries transformadas são característicos de séries não estacionárias com decaimento lento para zero. Observando os gráficos ACF e PACF, para uma e duas diferenças, analisou modelos ARIMA com uma e duas diferenças com o numero de lag significativos ACF para a médias móveis e PACF para a parte auto-regressiva. No qual identificou-se os modelos ARIMA(2,2,1) para a série com transformação logarítmica da quantidade de aves abatidas, o modelo ARIMA(4,2,1) para a série com transformação logarítmica da quantidade de bovinos abatidas, e o modelo ARIMA(3,2,1) para a série com transformação logarítmica da quantidade de suínos abatidas no qual ilustrou uma boa aderência nos modelos sugeridos.

**Palavras-chave:** *Séries transformadas, Séries não-estacionárias e Modelos ARIMA*

## 1 Introdução

A agropecuária desempenha um papel de grande importância no cenário da economia nacional, além disso, foi uma das primeiras atividades econômicas a serem desenvolvidas no país. A atividade agropecuária no país representa 8% do PIB (Produto Interno Bruto) brasileiro e gera

---

<sup>1</sup>LCE-ESALQ/USP. Email: [valiana@usp.br](mailto:valiana@usp.br)

<sup>2</sup>LCE-ESALQ/USP. Email: [djairdurand@usp.br](mailto:djairdurand@usp.br)

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

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

emprego para pelo menos 10% da população economicamente ativa do país. A produção agropecuária tem como objetivo destinar seus produtos, tais como grãos, frutas, verduras e também carne, leite, ovos, entre outros, para abastecer o mercado interno e especialmente o mercado externo.

O Brasil se destaca no cenário mundial na produção de carne de frango, bovina e suína, sendo as regiões sul e centro-oeste as principais regiões produtoras. Em 2013 o país bateu recordes em todos os segmentos, apresentando uma incrível marca de 34,4 milhões de cabeças bovinas, 36,1 milhões de cabeças de suínos e 5,6 bilhões de unidades de frango (IBGE).

O objetivo deste trabalho é utilizar modelos de séries temporal para fazer o estudo das séries de quantidades de aves, bovinos e suínos abatidos dos anos de 1975 a 2013 e sugerir modelos que possam ser úteis para explicar cada série.

## 2 Metodologia

Utilizou-se o conjunto de dados anuais referente a três séries univariadas, representada por quantidades de aves, bovinos e suínos abatidos do Brasil entre os anos de 1975 a 2013 (IPEA).

Para ajustar modelos a uma série temporal observada é necessário remover as fontes de variação não estacionária. Se a série observada for não estacionária pode-se tentar remover a tendência fazendo uma transformação, que consiste em tomar diferenças sucessivas da série original, até obter uma série estacionária. Este modelo é conhecido por Auto-regressivo integrado de média móvel (ARIMA).

O modelo ARIMA  $(p, d, q)$  é expresso por:

$$z_t = \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_{p+d} z_{t-p-d} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_q \varepsilon_{t-q}, \quad (1)$$

em que:  $z_t = y_t - y_{t-d}$ ;  $y_t$  é a observação no tempo  $t$ ;  $d$  é a diferença dos processos de combinação linear de valores passados;  $\theta_i$  é o parâmetro do modelo da parte auto-regressiva com  $i=1, \dots, p$ ;  $y_{t-i}$  é a observação no tempo  $t-i$ , com  $i=1, \dots, p$ ;  $\varepsilon_{t-j}$  é o erro no tempo  $t-j$ ;  $\varphi_j$  é o parâmetro do modelo da parte de médias móveis;  $\varepsilon_t$  representa o erro, supondo  $\varepsilon_t \sim N(0, \sigma^2)$ .

Para determinar a ordem adequada de um modelo é recomendado utilizar as funções de autocorrelação (ACF) e função autocorrelação parcial (PACF). As propriedades teóricas das funções de autocorrelação e autocorrelação parcial para alguns processos estacionários auxiliam na determinação da ordem  $(p, q)$  do modelo ARIMA, em que  $p$  é a ordem do modelo AR e  $q$  é a ordem do modelo MA (EHLERS, 2009). A tabela 1 apresenta as propriedades teóricas da ACF e PACF.

Tabela 1: Propriedades teóricas da ACF e PACF

Processo	ACF	PACF
Série aleatória	0	0
AR(1), $\theta > 0$	Decaimento exponencial	$0, k \geq 2$
AR(1), $\theta < 0$	Decaimento oscilatório	$0, k \geq 2$
AR( $p$ )	Decaimento para 0	$0, k > p$
MA(1)	$0, k > 1$	Decaimento oscilatório
ARMA( $p, q$ )	Decaimento a partir de $q$	Decaimento a partir de $p$

Um critério de seleção Critério de Informação de Akaike (AIC), é feito uma comparação entre dois modelos, e o modelo que tem o menor AIC é considerado o melhor modelo.

### 3 Resultados e discussões

Com base nas séries descritas anteriormente, apresenta-se a seguir os gráficos na Figura 1 no qual observa que a uma tendência de crescimento e um acréscimo na variância das séries de quantidades de aves, bovinos e suínos abatidos a medida que o tempo passa. Com isso uma transformação logarítmica pode ser adequada para estabilizar esta variância, assim passa se analisar o logaritmo de cada série.

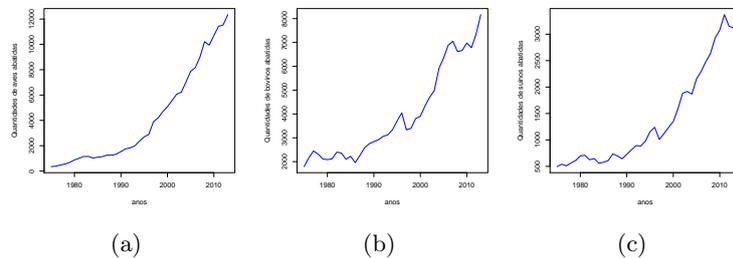


Figura 1: Gráfico das séries de aves (a), bovinos (b) e suínos (c) abatidos dos anos de 1975 a 2013

Construiu-se os gráficos das funções de autocorrelação (ACF) e de autocorrelação parcial (PACF) das séries com transformação logarítmica conforme mostrado na Figura 2. Observa-se que o comportamento do gráfico de ACF das séries de quantidades de aves, bovinos e suínos abatidos são característicos de séries não estacionarias com decaimento lento para zero.

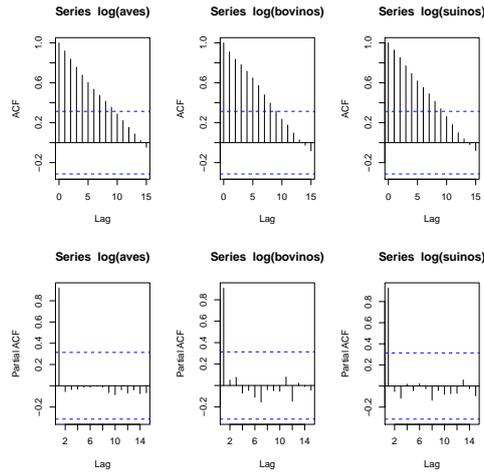


Figura 2: ACP e PACP das séries com transformação logarítmica da quantidade de aves, bovinos e suínos abatidas dos anos de 1975 a 2013

Então analisando o comportamento dos gráficos na Figura 3, com uma diferença, da ACF e PACF da série com transformação logarítmica da quantidades de aves abatidos não tem nenhum lag significativo . E o comportamento dos gráficos, com duas diferenças, da ACF da série com transformação logarítmica da quantidade de aves abatidas tem-se o lag 1 significativos e para a PACF o lag 1 e 2. O comportamento dos gráficos, com uma diferença, da ACF e da PACF da série com transformação logarítmica da quantidades de bovinos abatidos não tem nenhum lag significativo . E o comportamento dos gráficos, com duas diferenças, da ACF da série com transformação logarítmica da quantidade de bovinos abatidas tem-se o lag 1 significativo e para a PACF o lag 1 e 4. O comportamento dos gráficos, com uma diferença, da ACF e PACF da série com transformação logarítmica da quantidades de suínos abatidos tem o lag 15 significativo. E o comportamento dos gráficos, com duas diferenças, da ACF da série com transformação logarítmica da quantidade de suínos abatidas tem-se o lag 1 significativo e para a PACF o lag 1, 2 e o 3. Com isso verifica modelos ARIMA com uma e duas diferenças com o numero de lag significativos ACF para a médias móveis e PACF para a parte auto-regressiva.

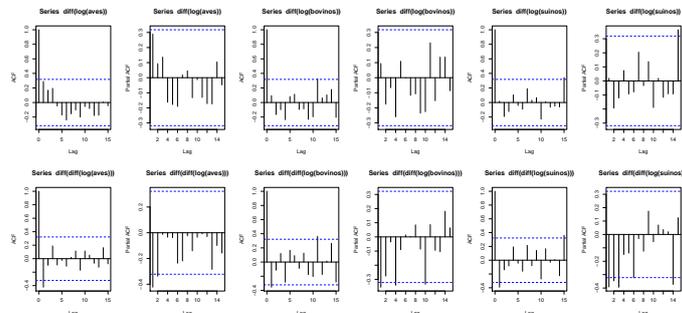


Figura 3: ACP e PACP das séries com transformação logarítmica, com uma e duas diferenças, de quantidade de aves, bovinos e suínos abatidas dos anos de 1975 a 2013

Após ajustar varias desses modelos, sugere o modelo que tem menor AIC e que os erros são ruídos brancos. No qual identificou-se os modelos ARIMA(2,2,1) para a série com transformação logarítmica da quantidade de aves abatidas, o modelo ARIMA(4,2,1) para a série com transformação logarítmica da quantidade de bovinos abatidas, e o modelo ARIMA(3,2,1) para a série com transformação logarítmica da quantidade de suínos abatidas no qual ilustrou uma boa aderência nos modelos apresentados na Figura 4.

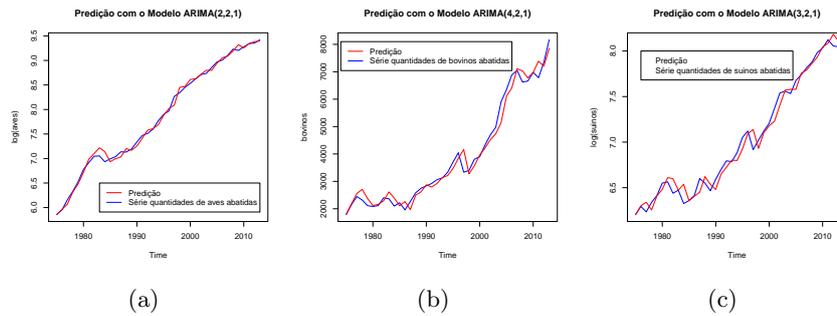


Figura 4: Predição das séries com transformação logarítmica da quantidade de aves (a), bovinos (b) e suínos (c) abatidas dos anos de 1975 a 2013

## 4 Conclusões

Com este trabalho, foi possível concluir que os modelos ajustados para predição da quantidade de aves, bovinos e suínos abatidos foram do tipo ARIMA e que foi preciso utilizar uma transformação logarítmica dos dados.

## Referências

- [1] EHLERS, R. S. **Análise de Séries Temporais**. 2009. Disponível em: <http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>
- [2] MORETTIN, P. A.; Toloi, C. M. C. **Análises de séries temporais**. São Paulo: Edgard Blucher. 2004.

# Aplicação de um teste *bootstrap* paramétrico para o índice de qualidade sensorial de berinjela minimamente processada

Rafael Lemos Bastos <sup>1 5</sup>

Eric Batista Ferreira <sup>2 5</sup>

Marcelo Silva de Oliveira <sup>3 5</sup>

Daniel Furtado Ferreira <sup>3 5</sup>

Ismael Simão <sup>4 5</sup>

*Na ciência dos alimentos, um dos campos mais importantes é a análise sensorial, onde avalia o produto e contribui para a determinação da sua qualidade e aceitação. Um ponto importante nesse campo é o estudo da vida de prateleira, pois a qualidade de um produto alimentício pode ser alterada no decorrer do tempo. Imm, Lee e Lee (2011) propuseram um Índice de Qualidade Sensorial (IQS) para indicar o comportamento da qualidade sensorial de um produto alimentício ao longo do tempo, de acordo com a reação do consumidor. Devido ao fato do IQS fornecer apenas uma estimativa pontual, sem realizar um teste de hipóteses de interesse com uma significância previamente estabelecida, Bastos (2013) propôs um teste bootstrap paramétrico (BP) para o IQS, tornando-o uma ferramenta melhor. O objetivo desse trabalho é apresentar o funcionamento do teste BP na seção dos métodos, estudar dados de berinjelas, aplicando esse teste neles. Foi utilizado o software R (R CORE TEAM, 2014) para aplicação do teste e representação gráfica. O funcionamento do teste BP foi apresentado em detalhes e aplicado nos dados de berinjela, informando que a qualidade sensorial das berinjelas avaliadas diminuiu, a 5% de significância. Portanto, os consumidores dificilmente ficarão satisfeitos com este alimento no fim da vida de prateleira.*

**Palavras-Chave:** Dados de Berinjelas, Método *bootstrap*, Software R.

---

<sup>1</sup>Graduado em Matemática Licenciatura, UNIFAL-MG; Mestre em Estatística e Experimentação Agropecuária, UFPA; Doutorando do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, UFPA. Lavras-MG. E-mail: [iel2702@yahoo.com.br](mailto:iel2702@yahoo.com.br)

<sup>2</sup>ICEX-UNIFAL-MG: Universidade Federal de Alfenas, Alfenas-MG

<sup>3</sup>DEX-UFPA: Universidade Federal de Lavras, Lavras-MG

<sup>4</sup>Graduado em Engenharia Agrícola, UFPA; Mestrando em Estatística e Experimentação Agropecuária, UFPA. Lavras-MG.

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

# **Análise e previsão da série temporal referente ao preço da cesta básica do município de São Paulo do período de janeiro de 1995 a novembro de 2013**

**Géssem de Faria Cesar<sup>1 5</sup>**

**Thaís de França Watanabe<sup>2 5</sup>**

**Luciene Resende Gonçalves<sup>3 5</sup>**

**Gabriel Rodrigo Gomes Pessanha<sup>4 5</sup>**

**Resumo:** Através da Lei nº 185 de 1936, estabeleceu-se que o salário mínimo é a remuneração destinada ao trabalhador, que satisfaça suas necessidades com habitação, vestuário, higiene, transporte e alimentação. Visto a deficiência de consumo calórico da população brasileira de baixa renda, a cesta básica busca prover, com base no valor do salário mínimo, uma lista de alimentos e suas quantidades preestabelecidas que assegurem a um indivíduo adulto quantidades ideais de proteínas, calorias, ferro, cálcio e fósforo. Em 1959, o DIEESE iniciou os cálculos do Índice de Custo de Vida no município de São Paulo passando, assim, a acompanhar o preço mensal da Cesta Básica Nacional. Esses cálculos são feitos com os preços coletados mensalmente dos produtos básicos e com suas quantidades estabelecidas no Decreto nº 399. A cesta básica é um agregado, que sofre reflexos dos fatores que influenciam os produtos que a compõem, o fator de maior impacto sobre a cesta é o preço. A série do preço da cesta básica do município de São Paulo, no período de janeiro de 1995 a novembro de 2013, sendo aplicada nesta os testes do Sinal e de Fisher, detectando a presença da componente tendência; em seguida aplicou-se a primeira diferença, tornando a série estacionária. Após ser considerada ruído branco, foram feitas quatro combinações, das quais a escolhida foi a que apresentou o menor Erro Quadrado Médio; assim, foi realizada a previsão para os quatro meses posteriores aos dados apresentados pela série.

**Palavras-chave:** *Cesta básica, Tendência, Estacionária, Combinação.*

---

<sup>1</sup> Graduando em Ciências Econômicas – Universidade Federal de Alfenas, gessem.cesar@gmail.com

<sup>2</sup> Graduanda em Ciências Econômicas – Universidade Federal de Alfenas, thais\_wat@hotmail.com

<sup>3</sup> Prof. Me. do ICSA – Universidade Federal de Alfenas, gabrielrgp@yahoo.com.br

<sup>4</sup> Profª Drª do ICSA – Universidade Federal de Alfenas, luciene.goncalves@unifal-mg.edu.br

<sup>5</sup> Agradecimentos á FAPEMIG pelo apoio financeiro

## **1. Introdução**

O salário mínimo é a remuneração destinada ao trabalhador, que satisfaça suas necessidades com habitação, vestuário, higiene, transporte e alimentação. No Brasil, como já se sabe, encontra-se nos extratos sociais de baixa renda deficiências de consumo calórico e de proteínas. A melhoria na alimentação dessa parcela mais pobre da população está intrinsecamente ligada aos preços dos alimentos básicos, que compõem a cesta básica.

A partir de um estudo censitário realizado em cada região do Brasil, e também de acordo com informações salariais extraídas das empresas regionais, criaram-se valores mínimos por região a serem pagos aos trabalhadores. Com isso, foi apresentada uma lista de alimentos com suas quantidades determinadas, a Cesta Básica Nacional, lista de alimentos com suas quantidades determinadas que fosse capaz de sustentar um trabalhador adulto assegurando quantidades ideais de proteínas, calorias, ferro, cálcio e fósforo.

O DIEESE, em janeiro de 1959, iniciou os cálculos do Índice de Custo de Vida no município de São Paulo, passando assim a acompanhar o preço mensal da Cesta Básica Nacional desde então. A cesta básica sofre reflexo direto dos fatores que influenciam os produtos que a compõem, sendo o preço o fator de maior importância. Esses cálculos são feitos com os preços coletados mensalmente dos produtos básicos, com suas quantidades específicas.

## **2. Objetivo**

O objetivo do presente trabalho é modelar a série da Cesta Básica do Município de São Paulo, a partir de janeiro de 1995 até novembro de 2013, visando deixá-la estacionária primeiramente para alcançar uma modelagem plausível e assim obter seus valores futuros.

## **3. Material e Métodos**

A série de dados foi coletada junto ao Ipeadata, com fonte do DIEESE. Os Dados consistem na série real dos Preços da Cesta Básica no município de São Paulo, no período de janeiro de 1959 a novembro de 2013. Porém neste trabalho foram utilizadas apenas 227 observações, coletadas entre o período de janeiro de 1995 a novembro de 2013, com o intuito de obter informações mais recentes.

Inicialmente foram construídos os gráficos da série, da função de autocorrelação, que são fundamentais na identificação das componentes tendência e sazonalidade. A literatura de Box e Jenkins recomenda que seja feita a identificação e o ajuste de diferentes modelos, antes de escolher um definitivo.

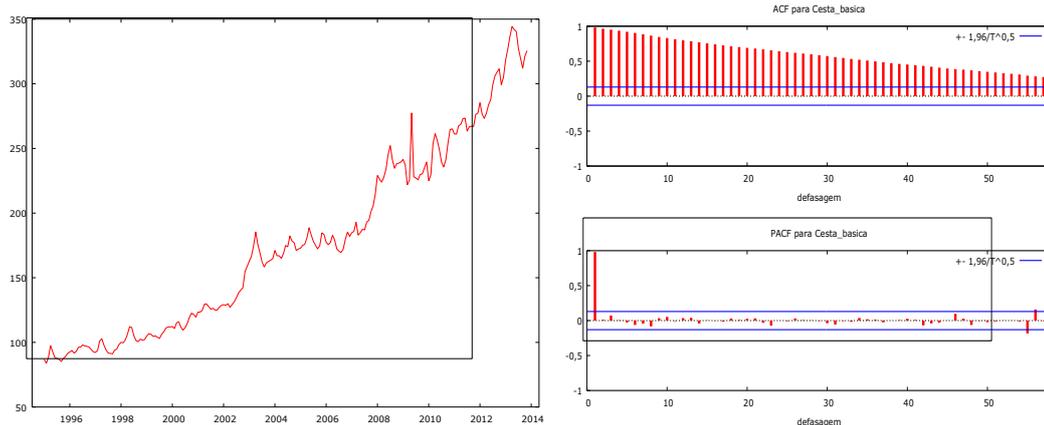


Figura1: Gráficos da série do preço da cesta básica no município de São Paulo, no período de janeiro de 1995 a novembro de 2013 e da sua função de autocorrelação.

Fonte: Dados do trabalho

O método utilizado no presente trabalho foi o de Box-Jenkins e o modelo foi o ARIMA. O método é utilizado para ajustar um modelo para o resíduo, por meio da série estacionária, sendo este escolhido através dos resultados obtidos na previsão, através do menor valor do Erro Quadrático Médio (EQMP). A Figura 1 apresenta os dados da série do preço da cesta básica no município de São Paulo e sua respectiva função de autocorrelação.

#### 4. Resultados e discussões

A Figura 2 apresenta a série sem tendência ou sazonalidade; a Figura 3 demonstra as funções de autocorrelação e de autocorrelação parcial.

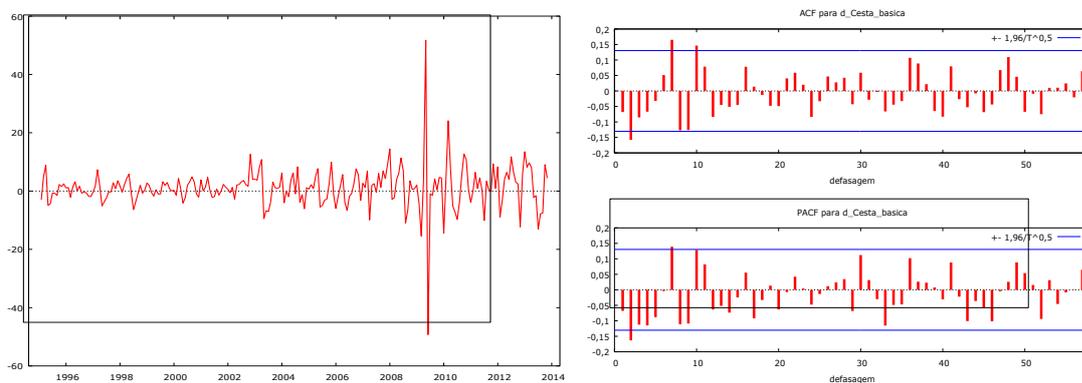


Figura 2: Gráficos do preço da cesta básica no município de São Paulo, no período de janeiro de 1995 a novembro de 2013, sem a presença das componentes tendência e sazonalidade, constituindo uma série estacionária e sua função de autocorrelação.

Fonte: Dados do trabalho

Com relação aos fatos apresentados, ajustaram-se quatro modelos, o ARIMA (0,1,2), em seguida um ARIMA (1,1,1), um ARIMA (1,1,2) e, por fim, um ARIMA (2,1,1). A seleção do melhor resultado tomou como base o Critério de Informação de Akaike (AIC) e do Erro Quadrado Médio (EQMP), como é apresentado na Tabela 1, o modelo que teve o melhor ajuste foi o ARIMA (1,1,1).

Tabela 1: Critério de Informação de Akaike (AIC) e no Erro Quadrado Médio (EQMP)

Modelos	AIC	EQMP
ARIMA (0,1,2)	1492,915	248,44
ARIMA (1,1,1)	1494,377	186,04
ARIMA (1,1,2)	1493,767	220,72
ARIMA (2,1,1)	1493,360	238,85

FONTE: Dados da pesquisa

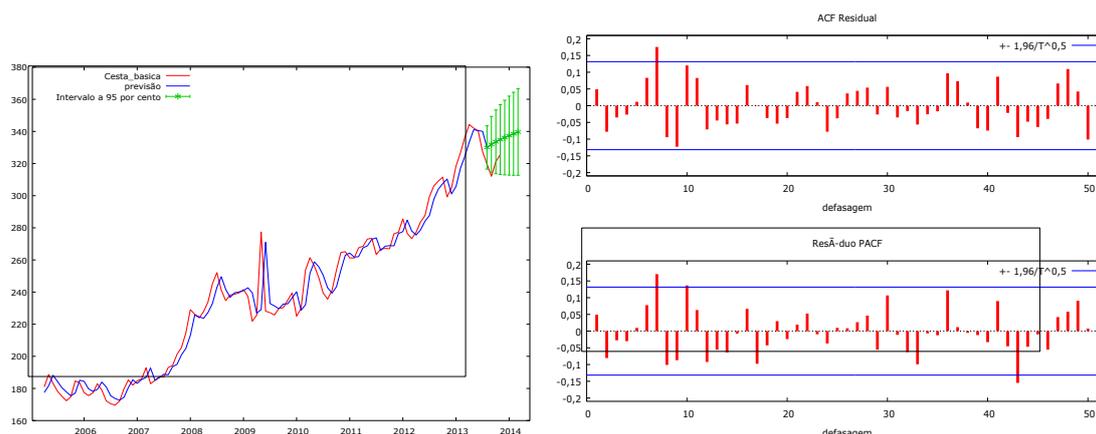


Figura 3: Gráfico da função de autocorrelação dos resíduos ajustados da série – ARIMA (1,1,1) e capacidade preditiva dos modelos ajustados.

Fonte: Dados do trabalho

O modelo que apresentou o menor valor do Critério de Akaike foi o ARIMA (0,1,2), porém este não foi escolhido devido a finalidade do presente trabalho, que é de realizar previsões. Portanto, o melhor modelo que ajustou a série foi o ARIMA (1,1,1), uma vez que apresentou o menor valor do Erro Quadrado Médio (EQMP). Os modelos estimados e suas capacidades foram apresentados na Figura 3, além da função de autocorrelação do modelo selecionado, o qual denota o comportamento de um ruído branco.

## 5. Conclusão

Os modelos de séries temporais foram utilizados para denotar o comportamento do preço da cesta básica no município de São Paulo; e a melhor combinação encontrada para esta foi a ARIMA (1,1,1), não havendo presença de intervenção. Assim, o objetivo do trabalho foi alcançado, uma vez que era o de realizar previsões de dados futuros, os quais não tiveram muita variância se comparados aos dados anteriores.

## REFERÊNCIAS BIBLIOGRÁFICAS

DIEESE. Acessado em: 18/01/2014. Disponível em: <http://www.dieese.org.br/>

GUJARATI, D. N.; PORTER, D.C. Econometria Básica, 5.ed., São Paulo: AMGH Editora LTDA, 2011.

IPEADATA. Preços da Cesta Básica do Município de São Paulo. Consultado em: 18/01/2014. Disponível em: [www.ipeadata.gov.com.br](http://www.ipeadata.gov.com.br)

MORETTIN, P. A. Análise de Séries Temporais, 3.ed., São Paulo: Egard Blucher, 2006.

# Modelo de mistura de distribuições aplicado a dados de atendimento de Procons

Rafael Agostinho Ferreira<sup>1 3</sup>

Deive Ciro de Oliveira<sup>2 3</sup>

**Resumo:** O Procon é um órgão que busca assegurar os direitos dos consumidores por meio do Código de Defesa do Consumidor. A demanda por estes serviços é registrada e publicada, periodicamente, pela Secretaria Nacional do Consumidor. Assim, o objetivo deste trabalho é analisar o comportamento, via mistura de distribuições, da frequência de ligações feitas aos Procons ao longo do primeiro semestre de 2012. Mistura de distribuições são usadas para modelar experimentos em que as observações de uma população possa ser composta por  $m$  subpopulações (componentes). Associada a cada uma delas está um peso  $\lambda_j$ , que representa a probabilidade da componente  $j$  pertencer à distribuição  $j$ . Para a estimação dos parâmetros do modelo, utiliza-se o algoritmo de iteração EM. Para este trabalho foi considerado dois modelos. Um considerando a mistura de distribuições Normais e outro mistura de Gamas, sendo este último o modelo mais verossímil (log verossimilhança de -4519789). A partir do modelo escolhido, foi possível calcular alguns valores esperados a cerca do comportamento dos acionamentos aos Procons. Logo, o número médio de atendimentos feitos pelos Procons, pela manhã, é de 2330 ligações. Já no período da tarde, esse número tende a cair, em média, para 1948 atendimentos. Portanto, é possível concluir que a demanda por atendentes é maior no período da manhã do que no período da tarde.

**Palavras-chave:** *Procon, ajuste, misturas de distribuições*

## 1 Introdução

A Fundação de Proteção e Defesa do Consumidor, o PROCON, é uma autarquia do Poder Executivo municipal ou estadual destinado à proteção e defesa dos direitos e interesses dos consumidores (CONSUMIDOR, 2014). Este órgão é responsável pelo contato mais direto com os cidadãos e suas questões, que busca assegurar que os direitos da sociedade sejam mantidos,

---

<sup>1</sup>Unifal-MG: Universidade Federal de Alfenas. Email: [rafaelferreiratc@hotmail.com](mailto:rafaelferreiratc@hotmail.com)

<sup>2</sup>Unifal-MG: Universidade Federal de Alfenas. Email: [deive.oliveira@unifal-mg.edu.br](mailto:deive.oliveira@unifal-mg.edu.br)

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

por meio da aplicação das normas do Código de Defesa do Consumidor. Para atingir tais fins, o PROCON utiliza de um sistema denominado SINDEC (Sistema Nacional de Informações de Defesa do Consumidor), o qual informatiza e cadastra os processos relacionados aos atendimentos prestados aos consumidores. O SINDEC possui um boletim, gerenciado pela Secretaria Nacional do Consumidor (SENACON), do qual publica e integra periodicamente, todos os atendimentos registrados ao PROCON (FEDERAL, 2014).

## 2 Objetivo

O presente trabalho busca analisar o comportamento, via mistura de distribuições, da frequência de ligações feitas aos PROCONs, ao longo do dia utilizando dados do primeiro semestre de 2012.

## 3 Dados e metodologia

### 3.1 Dados

Para a realização deste trabalho, foram coletados dados do número de ligações feitos aos PROCONs em todo o Brasil, referentes ao primeiro semestre do ano de 2012 no Portal Brasileiro de Dados Abertos (FEDERAL, 2014). Esses dados totalizaram 941660 observações e foram agrupados conforme a hora e data de ocorrência de cada ligação.

### 3.2 Metodologia

Mistura de distribuições são utilizadas para modelar experimentos em que as observações de uma população possa ser composta por  $m$  subpopulações (SARAIVA, 2009). Tal modelo, considerando distribuições independentes, consiste em um número finito  $m$  de componentes (distribuições),  $f_j(x_i)$ . Associado a cada uma delas está um peso  $\lambda_j$ , em que  $0 \leq \lambda_j \leq 1$  e  $\sum_{j=1}^m \lambda_j = 1$  (VARGAS, 2011). Portanto, a densidade de cada  $x_i$  pode ser escrita como:

$$g_{\Theta}(x_i) = \sum_{j=1}^m \lambda_j f_j(x_i | \theta_j)$$

em que  $\Theta = (\lambda, \theta) = (\lambda_1, \dots, \lambda_m, \theta_1, \dots, \theta_m)$  denotam os parâmetros da densidade de cada  $x_i$ .

Para a estimação dos parâmetros do modelo, foi utilizado um método numérico iterativo. Segundo Horta (2009), os parâmetros são estimados, através do algoritmo de iteração *EM*, por meio de dois passos (Passo-E e Passo-M). O Passo-E (*Expectation*) calcula a esperança da log verossimilhança ( $\log h_{\theta}(c) | X, \theta^{(t)}$ ) condicionado ao dado observado e valores iniciais definidos

aos parâmetros  $\theta$ . O Passo-M (*Maximization*) maximiza a log verossimilhança em função dos resultados obtidos no Passo-E.

## 4 Resultados e Discussões

Por meio da análise da Figura 1, é possível observar o comportamento da frequência de ligações feitas aos Procons ao longo do dia.

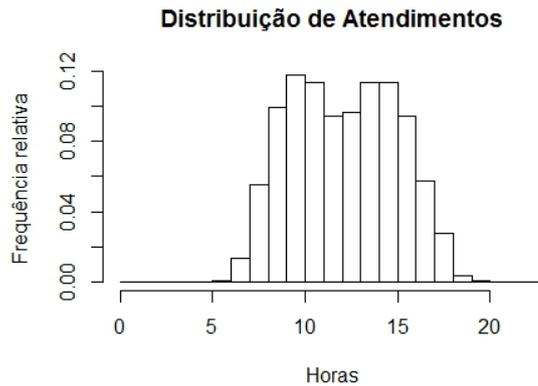


Figura 1: Histograma do número de ligações feitas aos Procons ao longo do dia

Para análise dos dados, utilizou-se o pacote *Mixtools* (BENAGLIA et al., 2009) do *software* R (TEAM, 2008). Esta biblioteca é utilizada para ajustes considerando misturas de distribuições. Para tal, foi considerado dois modelos. Um considerando mistura de distribuições Normais e outro mistura de Gammas. Assim, por meio do algoritmo *EM*, foi estimado os parâmetros de cada modelo.

Para a mistura de normais a densidade estimada foi:

$$g_{\hat{\Theta}}(x_i) = 0.4490799 \times f_1(x_i) + 0.5509201 \times f_2(x_i)$$

Onde:

- $f_1(x_i) = N(\hat{\mu}_1 = 10.02423, \hat{\sigma}_1^2 = 1.449729)$
- $f_2(x_i) = N(\hat{\mu}_2 = 14.67191, \hat{\sigma}_2^2 = 1.801987)$

Já para o segundo modelo, que considerou que as componentes seguiam distribuições Gama, apresentou como densidade estimada:

$$g_{\hat{\Theta}}(x_i) = 0.498407f_1(x_i) + 0.501593f_2(x_i)$$

Onde:

- $f_1(x_i) = \text{Gama}(\hat{\alpha}_1 = 49.7828053, \hat{\beta}_1 = 0.2010184)$
- $f_2(x_i) = \text{Gama}(\hat{\alpha}_2 = 68.1070630, \hat{\beta}_2 = 0.2154388)$

As densidades podem ser vistas na Figura 2(a) e 2(b):

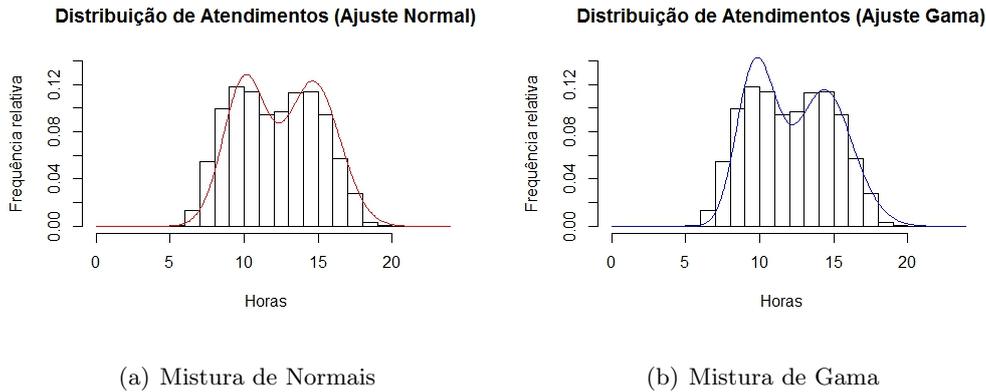


Figura 2: Ajuste de misturas para cada modelo

O modelo mais verossímil foi o de mistura de Gamas (log verossimilhança de -4519789). A partir do modelo escolhido, foi possível calcular alguns valores esperados a cerca do comportamento dos acionamentos aos Procons. Foram obtidos o número esperado de acionamentos diários no horário comercial. Para isso, foi calculado o número médio de ligações diárias no período, ponderadas pelas probabilidades de ocorrerem ligações no período da manhã (08 às 12 hs) e no período da tarde (13 às 17 hs), como é observado na Figura 3.

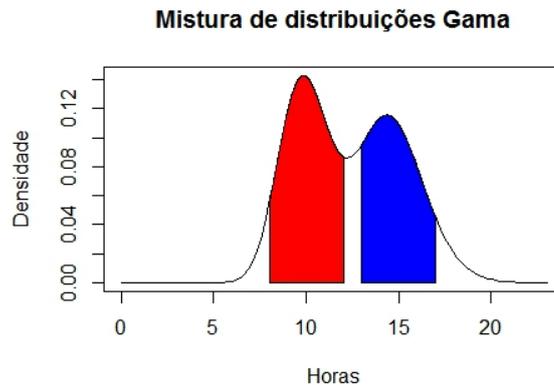


Figura 3: Região de integração para cálculo de probabilidade (vermelho [8 as 12]) (azul [13 as 17])

Logo, o número médio de atendimentos feitos pelos Procons, pela manhã, é de 2330 ligações. Já no período da tarde, esse número tende a cair, em média, para 1948 ligações.

## 5 Conclusões

Os Procons são órgãos de suma importância para a sociedade, pois este tem como objetivo defender os direitos dos consumidores frente às empresas prestadoras de bens e serviços. Assim, se faz necessário o conhecimento da frequência com que os consumidores demandam tais serviços, a fim de obter melhores práticas de gestão na prestação de serviços aos consumidores.

Portanto, este trabalho teve como objetivo modelar o comportamento de atendimentos dos Procons aos cidadãos. Assim, é possível concluir que a demanda por atendentes é maior no período da manhã do que no período da tarde.

## Referências

BENAGLIA, T. et al. mixtools: An r package for analyzing finite mixture models. *Journal of Statistical Software*, v. 32, n. 6, p. 1–29, 2009.

CONSUMIDOR, S. de estado de Proteção de Defesa do. <http://www.procon.rj.gov.br/>: [s.n.], 2014. Disponível em: <<http://www.procon.rj.gov.br/index.php/main/historico>>.

FEDERAL, G. *Atendimentos de Consumidores nos Procons (Sindec)*. Maio 2014. Disponível em: <<http://dados.gov.br/>>.

HORTA, M. M. *Modelos de mistura de distribuições na segmentação de imagens SAR polarimétricas multi-look*. Tese (Tese) — Universidade de São Paulo. Instituto de Física de São Carlos, 2009. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/76/76132/tde-06072009-215138/pt-br.php>>.

SARAIVA, E. F. *Modelo de Mistura com Número de Componentes Desconhecido: Estimacao via Método Split-Merge*. Tese (Doutorado) — Universidade Federal de Sao Carlos, 2009.

TEAM, R. D. C. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2008. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.

VARGAS, R. N. *Inferência estocástica e modelos de mistura de distribuições*. Tese (Dissertação) — Universidade Federal do Rio Grande do Sul. Instituto de Matemática. Programa de Pós-Graduação em Matemática., 2011. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/55330>>.

# Máxima Curvatura Modificada, Modelo de Regressão Linear com Platô e Modelo de Regressão Quadrático com Platô para a estimação de tamanho ótimo de parcelas em teste de germinação de sementes da espécie *Lafoensia pacari* A.

Carolina Juvelina Alves<sup>1 3</sup>  
Quintiliano S. S. Nomelini<sup>2 3</sup>

## Resumo

Com o intuito de reduzir a quantidade de sementes que serão testadas em testes de germinação, para diminuir as limitações de execução de um projeto pelas instituições de pesquisa e seu financiamento pelas agências de fomento o presente trabalho utiliza das metodologias: Método da Máxima Curvatura Modificado, Modelo de Regressão Linear com Platô, e Modelo de Regressão Quadrático com Platô, na espécie *Lafoensia pacari* A., em 3 lotes de variabilidades distintas para a determinação de tamanho ótimo de parcela da mesma. Considerou-se ensaio em branco e condições ambientais idênticas de todas as unidades experimentais básicas, com instalação dos experimentos de acordo com a metodologia proposta segundo Nomellini (2012). Verificou-se para esta espécie que os Modelos de Regressão (Linear e Platô) reduziram os tamanhos de parcela de forma drástica e prejudicial, sendo o menos prejudicial para a espécie o Método da Máxima Curvatura Modificado.

Palavras-chave: *Teste de Germinação, Tamanho Ótimo de Parcela, Lafoensia pacari* A.

## 1 Introdução

A destruição crescente e contínua da vegetação nativa visando à implantação de culturas agropastoris e a extração madeireira, sem a manutenção das áreas de reserva legal e proteção permanente e o desmatamento, deixando o solo descoberto, tem gerado, entre pesquisadores e a população grande discussão sobre os processos de conservação dos bens naturais. Consequentemente, preocupando pesquisadores de sementes e incentivando estudos sobre métodos de melhoria da qualidade de sementes florestais nativas.

Com o objetivo de controlar a variabilidade de resultados que ocorre entre as sementes cultivadas em 1931, a International Seed Testing Association (ISTA), determinou um tamanho máximo de lote para o controle da heterogeneidade. No entanto é notório que sementes de espécies florestais nativas apresentam maior variabilidade de resultados, quando comparadas às cultivadas, ou mesmo às florestais tropicais, pois como cada espécie possui características individuais. Assim um

---

<sup>1</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG. E-mail: caroljuvelina@est.ufu.br

<sup>2</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG. E-mail: quintiliano@famat.ufu.br

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

estudo mais específico sobre sementes florestais é muito importante, visto que estas possuem alta variabilidade, pois sofrem modificações genéticas naturais, diferentemente das cultivadas que ao longo do tempo são modificadas/selecionadas pelo homem visando maior homogeneidade e com alta qualidade.

Se a quantidade de sementes que será testado for muito grande, pode acarretar em maior custo do experimento, maior tempo para avaliação e gasto de material de consumo, o que pode limitar a execução de um projeto pelas instituições de pesquisa e seu financiamento pelas agências de fomento. Para se aumentar a precisão experimental e maximizar as informações obtidas no experimento a determinação do tamanho de parcela e do número de repetições, fornece uma maior precisão nos resultados obtidos (PARANAIBA, 2007).

Logo, justifica-se a determinação do número ótimo de parcelas para a espécie florestal *Lafoensia pacari* A. com o método de germinação validado por Nomelini(2012) utilizando-se dos Métodos da Máxima Curvatura Modificado (MMCM), Modelos de Regressão Quadrático com Platô (MRQP), Modelo de Regressão Linear com Platô (MRLP).

## 2 Objetivo

Estimar o tamanho ótimo de parcela para a espécie florestal *Lafoensia pacari* A. a ser utilizado em teste de germinação pelo método da Máxima Curvatura Modificado, de Modelos de Regressão Quadrático com Platô e de Regressão Linear com Platô. Comparar os resultados encontrados nos três métodos; Implementar os métodos no software livre R Development Core Team (2008).

## 3 Metodologia

Para a espécie florestal nativa *Lafoensia pacari* A. serão instalados os experimentos de acordo com a metodologia descrita na Tabela 1.1. Esta metodologia foi gerada a partir de testes e re-testes em laboratório, e é considerado como sendo o melhor para teste de germinação dessa espécie (NOMELINI, 2012).

**Tabela 1.1** Relação da espécie (família, nome popular e registro nacional de cultivar), com o método mais eficiente para teste de germinação incluindo substrato, temperatura, regime de luz, pré-tratamento, assepsia e contagens (NOMELINI, 2012).

Espécie (família) Nome popular/RNC	Metodologia	
<i>Lafoensia pacari</i> A. St.-Hil. (LYTHRACEAE) Pacari-verdadeiro 24305	Substrato	Rolo de Papel
	Temperatura/Luz	25°C/ contínua
	Pré-tratamento	-
	Assepsia	Solução de detergente
	Avaliação	1ª contagem
Final		21 dias

Para todas as técnicas serão considerados ensaio em branco, condições ambientais idênticas de todas as unidades experimentais básicas (UEB), e os tratamentos serão três lotes de sementes de qualidades fisiológicas distintas, os quais formaram lotes de alta, média e baixa porcentagem de germinação. Os dados foram coletados em um gride  $r \times c$ , com  $r$  linhas e  $c$  colunas, com um total de

$rc$  UEB, onde  $r$  representa os laboratórios e  $c$  as repetições, Figura 1.1. Os laboratórios escolhidos foram considerados idênticos em estudo feito por Nomelini (2012).

$r_1c_1$	$r_1c_2$	$\cdots$	$r_1c_n$
$r_2c_1$	$r_2c_2$	$\cdots$	$r_2c_n$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$r_nc_1$	$r_nc_2$	$\cdots$	$r_nc_n$

**Figura 1.1** Croqui de um ensaio em branco.

Foram formados vários tamanhos de parcelas com  $X$  UEB agrupadas de modo conveniente, em que  $X$  refere-se a um divisor de  $rc$ . Para cada valor de  $X$ , há um coeficiente de variação associado calculado segundo Paranaíba(2007), onde  $CV(x) = (\sqrt{Var(x)}/\bar{y}(x)) * 100$ , em que  $CV(x)$  é o coeficiente de variação entre parcelas de tamanhos ( $X$ );  $Var(x)$  é a variância entre as parcelas de  $X$  UEB;  $\bar{y}(x)$  é a média das parcelas de  $X$  UEB. Para o MMCM, a relação entre o coeficiente de variação ( $CV(x)$ ) e o correspondente tamanho da parcela com  $X$  UEB será estimada pelo modelo  $CV(x) = a/x^b$  em que  $a$  e  $b$  são os parâmetros a serem estimados. O valor da abscissa em que ocorre o ponto de máxima curvatura  $x_c$  pode ser estimado por:  $x_c = \exp\{[1/(2b+2)]\log[(ab)^2(2b+1)/(b+2)]\}$  em que  $x_c$  é o valor da abscissa no ponto de máxima curvatura, o qual corresponde à estimativa do tamanho ótimo da parcela experimental;  $a$  é a estimativa da constante de regressão; e  $b$  é o coeficiente de regressão.

Nos MRLP e MRQP, serão utilizados os seguintes modelos, respectivamente:

$$CV_i = \begin{cases} \beta_0 + \beta_1 X_i + \varepsilon_i \rightarrow x_i \leq x_c \\ P + \varepsilon_i \rightarrow x_i > x_c \end{cases}, CV_i = \begin{cases} \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i \rightarrow x_i \leq x_c \\ P + \varepsilon_i \rightarrow x_i > x_c \end{cases}$$

em que  $CV_i$  é o coeficiente de variação entre as repetições de cada tamanho de parcela;  $X_i$  é o tamanho da parcela em UEB;  $x_c$  é o tamanho ótimo de parcelas para o qual o modelo linear ou quadrático se transforma em um platô;  $P$  é o coeficiente de variação no ponto correspondente ao platô;  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  são os coeficientes a serem estimados,  $\varepsilon_i$  é o erro associado ao  $CV_i$  considerados independentes e normalmente distribuídos, com média zero e variância constante. Para o MRLP o tamanho ótimo de parcela será estimado pela expressão:  $\hat{x}_c = \hat{P} - \hat{\beta}_0/\hat{\beta}_1$ , onde  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  e  $\hat{P}$  são os valores estimados dos parâmetros do modelo. Para o MRQP, o tamanho ótimo de parcela e a junção com o modelo platô serão estimados, respectivamente por:  $\hat{x}_c = -(\hat{\beta}_1/2\hat{\beta}_2)$ ,  $\hat{P} = \hat{\beta}_0 - (\hat{\beta}_1^2/4\hat{\beta}_2)$ .

Para todos os métodos, estimar-se-à os coeficientes de determinação para verificar a qualidade de ajuste dos modelos, dado por:  $R^2 = 1 - QM_E/QM_T$ , em que,  $QM_E$  é o quadrado médio do erro ou resíduo e  $QM_T$  é o quadrado médio total corrigido.

## 4 Resultados e Discussões

Para os três métodos foram utilizado os mesmos ensaios em branco, montados como na Figura 1.1, tendo 6 laboratórios, onde para cada laboratório foram enviadas 200 sementes e realizadas 4 repetições em cada lote, onde cada tratamento representa um lote sendo respectivamente lotes de alta,

média e baixa variabilidade, totalizando em cada um deles 24 UEB, as quais podem ser agrupadas de 6 modos diferentes com seus respectivos  $CV(x)_s$ , conforme a Tabela 1.2.

**Tabela 1.2.** Tamanho de parcela ( $x$ ), em termos de unidades experimentais básicas (UEB) e coeficiente de variação entre parcelas ( $CV(x)$ ), em % para o Lote 1, Lote 2 e Lote 3.

UEB	Tamanho da parcela( $x$ )	$CV(x)$ Lote 1	$CV(x)$ Lote 2	$CV(x)$ Lote 3
24	1	5,9885	12,7349	16,5863
12	2	4,5369	9,0955	10,9144
6	4	3,2896	7,8224	8,4686
4	6	1,6200	5,6092	5,3453
3	8	1,4396	7,3572	2,5050
2	12	1,1006	6,1636	1,7532

Utilizando-se os valores da Tabela 1.3, o software R e os parâmetros  $a$  e  $b$  da equação  $CV(x) = a/x^b$ , os resultados obtidos para os 3 lotes no MMCM, constam na Tabela 1.3.

**Tabela 1.3.** Lote 1 – alta variabilidade, Lote 2 – média variabilidade, Lote 3 – baixa variabilidade, Constante de regressão ( $\hat{a}$ ), Coeficiente de regressão ( $\hat{\beta}$ ), Tamanho da Parcela ( $\hat{x}_c$ ), Coeficiente de determinação ( $\hat{R}^2$ ) para o MMCM.

	Lote 1	Lote 2	Lote 3
$\hat{a}$	6,2640	12,1905	17,0365
$\hat{\beta}$	0,6171	0,3152	0,6888
$\hat{R}^2$	2,1973	2,4356	4,1444
$\hat{x}_c$	0,9334	0,8507	0,9341

Para o Lote 1 temos aproximadamente 2 parcelas estimadas pelo método MMCM, dividindo 24 por 2 parcelas temos aproximadamente 12 UEB, distribuídos entre os 6 laboratórios tem-se então 1 UEB por laboratório. Como todos os UEB são constituídas de 25 sementes, cada laboratório receberá 25 sementes do Lote 1; analogamente no Lote 2 teremos aproximadamente 12 UEB, 1 UEB/laboratório, no total de 25 sementes; para o Lote 3 teremos aproximadamente 6 UEB, 1 UEB/laboratório com total de 25 sementes. Então, o total de sementes enviadas para cada laboratório será de 125, bem inferior às 600 sementes enviadas inicialmente, uma redução de 79,67%.

Utilizando-se os valores da Tabela 1.3, o software R e os parâmetros  $\beta_0$ ,  $\beta_1$ ,  $P$ , da equação  $CV_i = \begin{cases} \beta_0 + \beta_1 X_i + \varepsilon_i \rightarrow x_i \leq x_c \\ P + \varepsilon_i \rightarrow x_i > x_c \end{cases}$ , os resultados obtidos para os 3 lotes no MRLP, constam na Tabela 1.4.

**Tabela 1.4.** Lote 1 – alta variabilidade, Lote 2 – média variabilidade, Lote 3 – baixa variabilidade, Parâmetros do modelo ( $\hat{\beta}_0$  e  $\hat{\beta}_1$ ), Coeficiente de variação ( $\hat{P}$ ), Tamanho da Parcela ( $\hat{x}_c$ ), Coeficiente de determinação ( $\hat{R}^2$ ) para o MRLP.

	Lote 1	Lote 2	Lote 3
$\hat{\beta}_0$	6,5520	13,3715	16,9192
$\hat{\beta}_1$	-0,8287	-1,4945	-2,0279
$\hat{R}^2$	0,9776	0,7897	0,9216
$\hat{P}$	1,2701	6,3767	2,1291
$\hat{x}_c$	6,3738	4,6803	7,2934

Analogamente ao cálculo feito no MMCM, no Lote 1 teremos aproximadamente 4 UEB, 1 UEB/laboratório, no total de 25 sementes; no Lote 2 teremos aproximadamente 5 UEB, 1 UEB/laboratório, no total de 25 sementes; para o Lote 3 teremos aproximadamente 4 UEB, 1

UEB/laboratório no total de 25 sementes. Então, o total de sementes enviadas será de 75, bem inferior às 600 sementes enviadas inicialmente, uma redução de 87,5%.

Utilizando-se os valores da Tabela 1.3, o software R e os parâmetros  $\beta_0, \beta_1, \beta_2, P$ , da equação  $CV_i = \begin{cases} \beta_0 + \beta_1 X_i + \varepsilon_i \rightarrow x_i \leq x_c \\ P + \varepsilon_i \rightarrow x_i > x_c \end{cases}$ , os resultados obtidos para os 3 lotes no MRQP, constam na Tabela 1.5.

**Tabela 1.5.** Lote 1 – alta variabilidade, Lote 2 – média variabilidade, Lote 3 – baixa variabilidade, Parâmetros do modelo ( $\hat{\beta}_0, \hat{\beta}_1$  e  $\hat{\beta}_2$ ), Coeficiente de variação ( $\hat{P}$ ), Tamanho da Parcela ( $\hat{x}_c$ ), Coeficiente de determinação ( $\hat{R}^2$ ) para o MRQP

	<b>Lote 1</b>	<b>Lote 2</b>	<b>Lote 3</b>
$\hat{\beta}_0$	7,1121	15,0990	18,4742
$\hat{\beta}_1$	-1,3105	-3,1126	-3,1991
$\hat{\beta}_2$	0,0724	0,2804	0,1535
$\hat{R}^2$	0,9798	0,8426	0,9478
$\hat{P}$	1,1789	6,4618	1,8079
$\hat{x}_c$	9,0550	5,5498	10,4193

Analogamente ao cálculo feito no MMCM, no Lote 1 teremos aproximadamente 3 UEB, 1 UEB/laboratório, no total de 25 sementes; no Lote 2 teremos aproximadamente 4 UEB, 1 UEB/laboratório, no total de 25 sementes; para o Lote 3 teremos aproximadamente 3 UEB, 1 UEB/laboratório no total de 25 sementes. Então, o total de sementes enviadas será de 75, bem inferior às 600 sementes enviadas inicialmente, uma redução de 87,5%.

O  $\hat{R}^2$  nos três lotes indica um bom ajuste dos dados à equação em todos os modelos.

## 5 Conclusões

Os métodos Modelo de Regressão Linear com Platô e Modelo de Regressão Quadrático com Platô mostraram-se inadequados para a espécie *Lafoensia pacari* A., visto que obtiveram uma redução drástica nos tamanhos de parcelas nos 3 lotes analisados. O Método da Máxima Curvatura Modificado apresentou uma menor redução nestes tamanhos de parcelas sendo então entre estes o mais adequado para a espécie. Assim é recomendável o uso de mais de um método de determinação de tamanho ótimo de parcelas experimentais, visto que em cada espécie podemos obter resultados distintos.

## Referências

- NOMELINI, Q. S. S. *Enfoque estatístico na validação de métodos para teste de germinação de sementes de florestais*. 163p. Tese (Doutorado em Agronomia/Fitotecnia) – Universidade Federal de Uberlândia, Uberlândia. 2012.
- PARANAÍBA, P. F. *Tamanho ótimo de parcelas em delineamentos experimentais*. 63 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG. 2007.
- R Development Core Team. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>. (2008).

# Análise e previsão do rendimento da caderneta de poupança no período de agosto de 1994 e abril de 2014 utilizando modelos de séries temporais

Thaís de França Watanabe<sup>1 4</sup>  
Gabriel Rodrigo Gomes Pessanha<sup>2 4</sup>  
Géssem de Faria Cesar<sup>3 4</sup>

**Resumo:** A poupança é considerada a modalidade de investimento mais tradicional no Brasil, sendo classificada como conservadora por oferecer baixo risco, e conseqüentemente menor retorno, se comparado com outros tipos de aplicações financeiras. Desde seu início em 1861 até hoje, muitas modificações foram feitas com relação a rentabilidade da caderneta de poupança no Brasil. Nos dias atuais, com base na Medida Provisória N° 567 de 2012, a rentabilidade passou a depender da meta da Taxa SELIC, se essa atingir níveis superiores a 8,5% ao ano, a remuneração será de 0,5% ao mês, acrescido a TR; porém, se a meta da SELIC for inferior a 8,5% ao ano, a remuneração será de 70% da SELIC acrescido a TR. A série do rendimento da caderneta de poupança no primeiro dia útil do mês, no período de agosto de 1994 e abril de 2014, foi deflacionada através do deflator IGP-DI, e após a aplicação dos teste do Sinal e de Fisher, foi detectada a presença da componente tendência; em seguida através da aplicação da primeira diferença, a mesma tornou-se estacionária. Após a série ser considerada um ruído branco, foram propostas quatro combinações, das quais foi escolhida a que apresentou o menor Erro Quadrado Médio e portanto a combinação que melhor representou a série; a partir disso, foi feita uma comparação entre os dados previstos e os dados da série real, durante o período de julho de 2013 e abril de 2014, o qual havia sido reservado.

**Palavras-chave:** *Poupança, Tendência, Estacionária, Combinação.*

## 1. Introdução

A caderneta de poupança, criada no Brasil em meados do século XIX, tem como objetivo recolher depósitos, principalmente das classes menos favorecidas, para que essas pudessem poupar reservas para o futuro a fim de garantir um investimento mais seguro e com menos riscos.

A poupança é a forma de investimento mais segura e tradicional do Brasil, envolve baixo risco. Garante através do Fundo Garantidor de Crédito, atualmente, uma quantia máxima a ser

---

<sup>1</sup> Graduanda em Ciências Econômicas – Universidade Federal de Alfenas, thais\_wat@hotmail.com

<sup>2</sup> Prof.º Me. do ICSA – Universidade Federal de Alfenas, gabrielrgp@yahoo.com.br

<sup>3</sup> Graduando em Ciências Econômicas – Universidade Federal de Alfenas, gessem.cesar@gmail.com

<sup>4</sup> Agradecimentos pelo auxílio financeira à FAPEMIG

restituída de R\$ 250 mil em caso de falência, intervenção ou liquidação da instituição financeira em que se tem seu dinheiro aplicado. A garantia de baixo risco vem acompanhada também da baixa rentabilidade proporcionada pela caderneta, a menos rentável dentre as aplicações do mercado financeiro.

Desde seu início em 1861 até hoje, muitas modificações foram feitas com relação a rentabilidade da caderneta de poupança no Brasil. Nos dias atuais, com base na Medida Provisória N° 567 de 2012, a rentabilidade passou a depender da meta da Taxa SELIC, se essa atingir níveis superiores a 8,5% ao ano, a remuneração será de 0,5% ao mês, acrescido a TR; porém, se a meta da SELIC for inferior a 8,5% ao ano, a remuneração será de 70% da SELIC acrescido a TR.

## **2. Objetivo**

Sendo assim, o objetivo central deste trabalho é construir modelos estatísticos de previsão para série do rendimento mensal da caderneta de poupança no primeiro dia útil do mês, utilizando a metodologia de Box-Jenkins. Para tanto, reservou-se o período de julho de 2013 até o mês de abril de 2014 para a realização de comparações. Assim, após a estimação dos modelos e estes apresentarem seus resíduos em forma de ruído branco procederá a escolha do modelo que apresentar o melhor ajuste.

## **3. Material e Métodos**

A série utilizada para análise é a série do rendimento nominal da caderneta de poupança no primeiro dia útil do mês, compreendida entre agosto de 1994 e abril de 2014. Os dados foram obtidos junto ao IPEA – data (Instituto de Pesquisa Econômica Aplicada).

Para a identificação dos modelos de Box e Jenkins (contexto univariado), analisa-se as funções de autocorrelação e autocorrelação parcial da série em questão. No processo de identificação determina-se os valores de  $p$ ,  $d$  e  $q$  do modelo ARIMA e a ordem dos polinômios  $\omega_j$  e  $\delta_j$  e da constante de defasagem dos modelos de função de transferência. A literatura que trata da metodologia de Box e Jenkins recomenda a identificação de vários modelos experimentais que deverão ser ajustados e verificados, antes da escolha do modelo definitivo.

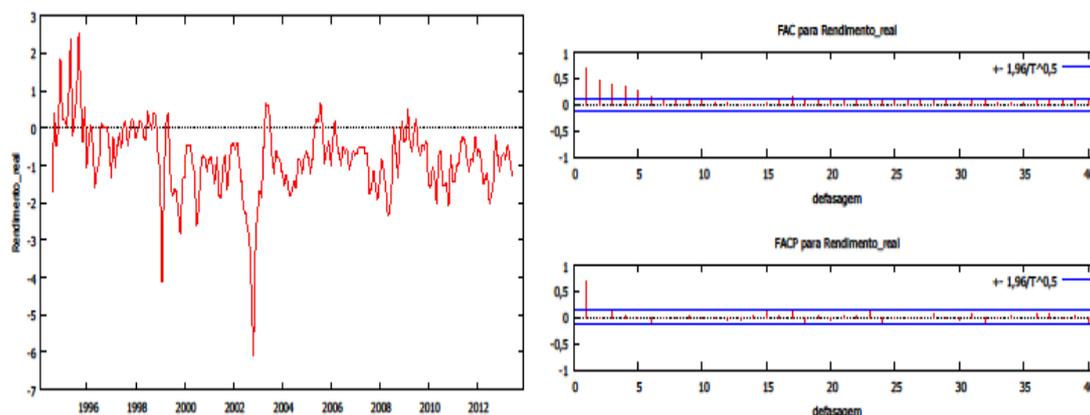


FIGURA 1: Gráficos da série do rendimento real da caderneta de poupança no primeiro dia útil do mês, no período de agosto de 1994 a junho de 2013 e da sua respectiva função de autocorrelação.

Fonte: Dados do Trabalho

O método utilizado neste trabalho foi de Box & Jenkins e o modelo foi o ARIMA. O método é empregado para ajustar um modelo para o resíduo por meio da série estacionária, sendo esse escolhido de acordo com os resultados obtidos pela previsão, pelo erro quadrático

#### 4. Resultados e Discussões

A série sem tendência e sem sazonalidade é representada na Figura 2. As funções de autocorrelação e de autocorrelação parcial são demonstradas através da Figura 3.

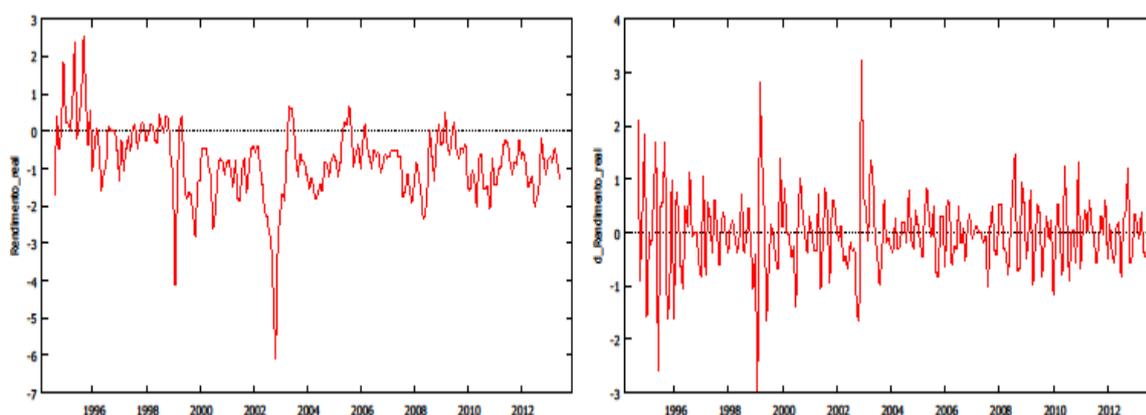


FIGURA 2: Gráfico do rendimento real da caderneta de poupança no primeiro dia útil do mês, no período de agosto de 1994 a junho de 2013, sem a presença de tendência e sazonalidade, constituindo uma série estacionária.

Fonte: Dados do Trabalho

A constatação de um bom ajustamento de um modelo se dá através da presença de um resíduo que constitua um ruído branco. Com base, na função de autocorrelação da série (Figura 3) pode-se

observar que a série ainda apresenta correlação entre os resíduos. Para este resíduo ajustou-se um modelo de Box & Jenkins.

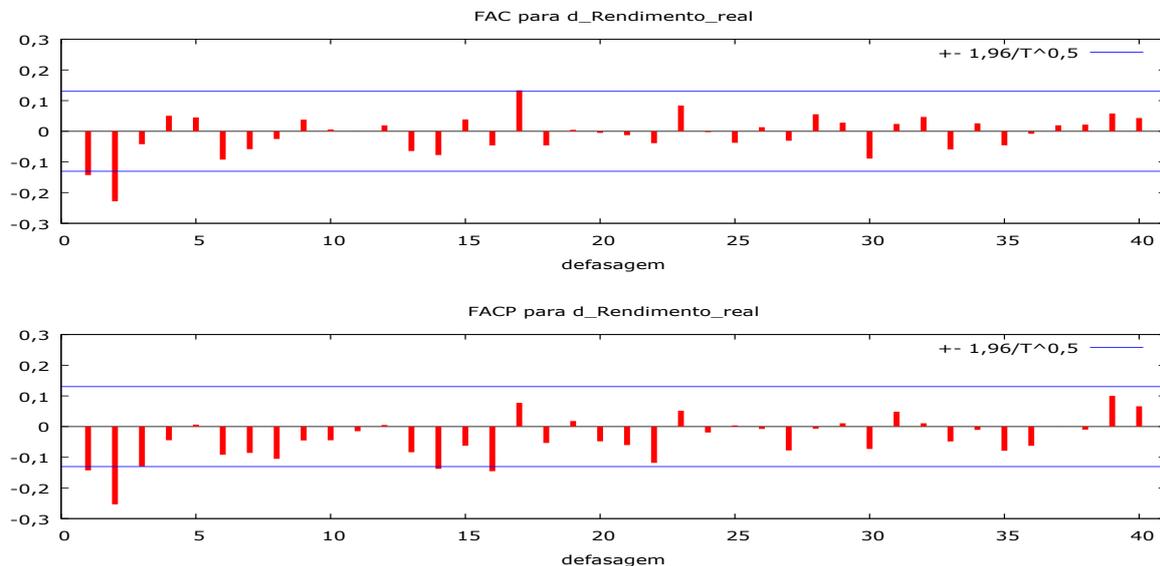


FIGURA 3: Função de autocorrelação e de autocorrelação parcial da série do rendimento real da caderneta de poupança no primeiro dia útil do mês, no período de agosto 1994 a junho de 2013, sem a presença de tendência e sazonalidade, constituindo uma série estacionária.

Fonte: Dados do Trabalho

Com base neste fato, ajustaram-se quatro modelos para posterior comparação da capacidade preditiva dos mesmos. Inicialmente, ajustou-se um modelo ARIMA (1,1,1), em seguida um ARIMA (3,1,1), um ARIMA (2,1,2) e, finalmente um ARIMA (0,1,3). A seleção do melhor resultado foi baseada no Critério de Informação de Akaike (AIC) e no Erro Quadrado Médio (EQMP), como mostra a Tabela 1, o modelo que apresentou melhor ajuste foi o ARIMA (3,1,1).

TABELA 1: Critério de Informação de Akaike (AIC) e no Erro Quadrado Médio (EQMP)

Modelos	AIC	EQMP
ARIMA (1,1,1)	485,0855	0,19048
ARIMA (3,1,1)	485,0709	0,18487
ARIMA (2,1,2)	484,6964	0,20402
ARIMA (0,1,3)	492,2459	0,20739

FONTE: Dados da pesquisa

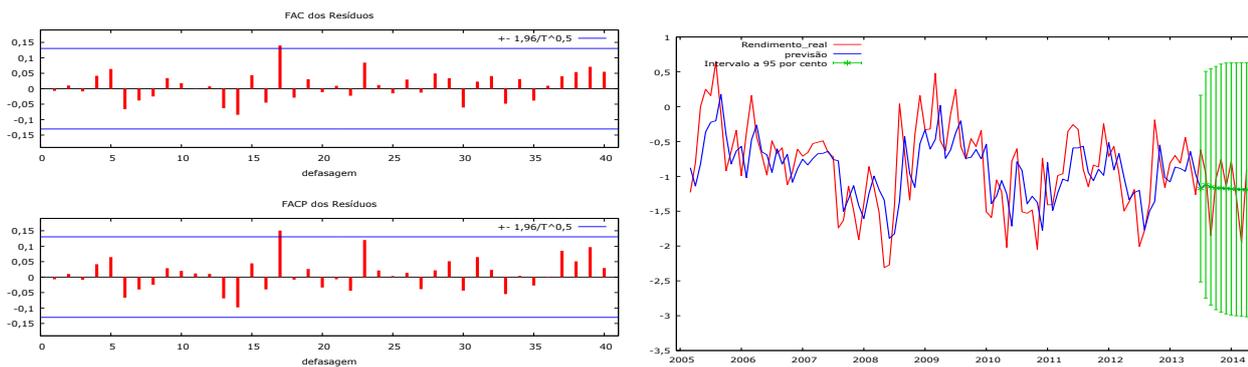


FIGURA 4: Gráfico da função de autocorrelação dos resíduos ajustados da série – ARIMA (3,1,1) e capacidade preditiva dos modelos ajustados.

Fonte: Dados do Trabalho

O modelo que apresentou o menor valor referente ao critério de Akaike foi o ARIMA com combinação (2,1,2). Porém, ele não foi escolhido como a melhor modelagem devido a finalidade do presente trabalho ser fazer previsões. O modelo que melhor se ajustou a série foi o ARIMA com combinação (3,1,1) pois apresentou o menor erro quadrado médio (EQMP) em detrimento aos outras combinações testadas. Os modelos estimados e suas respectivas capacidades preditiva estão representados na Figura 4, além disso, apresenta-se também na Figura 4 a função de autocorrelação do modelo selecionado que denota o comportamento do resíduo como sendo um ruído branco.

## 5. Conclusão

Os modelos de séries temporais podem ser utilizados para descrever a série do rendimento da poupança no primeiro dia útil do mês. Para tal série, o melhor modelo encontrado foi o ARIMA (3,1,1) sem a ocorrência de intervenções. Mesmo assim o modelo apresentou previsões próximas aos valores verdadeiros observados no período de julho de 2013 a abril de 2014.

## REFERÊNCIAS BIBLIOGRÁFICAS

IPEA DATA, 2006. **Instituto de Pesquisa Econômica Aplicada**. Disponível em: <<http://www.ipeadata.gov.br>>. Acesso: 15/02/2014.

MORRETIN, P. A.; TOLOI, C. M. C. **Previsão de séries temporais**. 2. ed. São Paulo: Atual Editora, 1987.

# Caracterização dos casos de atendimentos anti-rábicos humanos na cidade de Maringá-PR

Michelle Aparecida Corrêa <sup>1 5</sup>

Natalia da Silva Martins <sup>2 5</sup>

Elcio do Nascimento Chagas <sup>3 5</sup>

Eric Batista Ferreira <sup>4 5</sup>

A raiva humana, uma doença que é causada por um vírus, que apresenta como principal característica comprometer o sistema nervoso central sob a forma de encefalite. Sua transmissão se dá por secreções infectadas, usualmente saliva, as quais atingem o organismo por meio da mordedura animal ou da ferida aberta. A cidade de Maringá, que está situada no estado do Paraná, destaca-se pela qualidade de vida oferecida à seus moradores e alto IDH, por meio da 15ª Regional de Saúde realizou um levantamento em relação aos casos de raiva humana. E, por meio deste levantamento, apresentar uma caracterização dos casos de atendimentos anti-rábicos humanos, ocorridos no ano de 2012, avaliando possíveis associação entre as variáveis. Verificou-se que o perfil dos pacientes atendidos são predominantemente do sexo masculino, cor branca, possuem ensino médio completo e residem na zona urbana de Maringá. Os acidentes mais comuns são as mordeduras de cães, com ferimentos únicos e superficiais, sendo as mãos e pés as partes do corpo mais atingidas.

**Palavras-chave:** *atendimento anti-rábicos, raiva humana, Maringá.*

## 1 Introdução

A raiva humana, é uma doença causada por um vírus, que tem como principal característica comprometer o sistema nervoso central, sob a forma de encefalite. Sua transmissão se dá por secreções infectadas, usualmente saliva, as quais atingem o organismo por meio da mordedura animal ou da ferida aberta. Sua importância deriva do fato de ser letal em aproximadamente 100% dos casos (BABBONI; MODOLO, 2011). Segundo Silva et. al. (2009), no ciclo urbano, as principais fontes de infecção são o cão e o gato. No Brasil, o morcego é o principal responsável pela manutenção da cadeia silvestre. A Figura 1 apresenta os ciclos epidemiológicos da transmissão.

<sup>1</sup>ICEx - UNIFAL-MG. E-mail: *michellecorrea2007@hotmail.com.*

<sup>2</sup>ICEx - UNIFAL-MG. E-mail: *natalia.martins@unifal-mg.edu.br*

<sup>3</sup>IFES- Campus de Alegre. E-mail: *enchagas@ifes.edu.br*

<sup>4</sup>ICEx - UNIFAL-MG. E-mail: *eric.ferreira@unifal-mg.edu.br*

<sup>5</sup>Agradecimento à FAPEMIG e à CAPES pelo apoio financeiro.

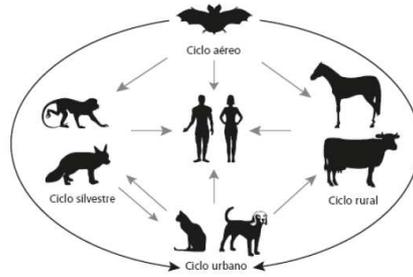


Figura 1: Ciclos epidemiológicos da transmissão da raiva

Devido ao elevado número de casos de raiva humana transmitida principalmente por cães nas décadas de 1950 e 1960 no Brasil, municípios e estados desenvolveram atividades e regulamentações direcionadas ao controle de zoonoses, em particular da raiva (WADA; ROCHA; MAIA-ELKHOURY, 2011). Deste modo, a cidade de Maringá, situada no estado do Paraná ( $23^{\circ} 25' 30''$  S,  $51^{\circ} 56' 20''$  W), a qual destaca-se pela qualidade de vida oferecida à seus moradores e alto Índice de Desenvolvimento Humano (IDH), por meio da 15<sup>a</sup> Regional de Saúde realizou um levantamento em relação aos casos de raiva humana. E, por meio deste levantamento, atender os objetivos propostos, os quais serão apresentados na seção ??.

## 2 Objetivos

Os objetivos deste estudo são: (i) Apresentar uma caracterização dos casos de atendimentos anti-rábiticos humanos na cidade de Maringá, ocorridos no ano de 2012; (ii) Avaliar possíveis associação entre as variáveis tipo de animal, tipo de ferimento, número de ferimentos e localização do ferimento.

## 3 Materiais e métodos

Os dados utilizados para o desenvolvimento deste trabalho são provenientes de um estudo sobre 1636 registros de casos de atendimentos anti-rábiticos da cidade de Maringá-PR, no ano de 2012. Para a coleta dos dados, consideraram-se as fichas de atendimento anti-rábitico humano do Sistema de Informação de Agravos de Notificação (SINAN), obtidas junto à 15<sup>a</sup> Regional de Saúde. As fichas de atendimentos foram utilizados para a presente estudo apresentando as seguintes variáveis: Mês da ocorrência, idade e gênero, tipo de exposição, região atingida, características do ferimento, quanto à extensão e número de lesões, e quanto à profundidade, espécie do animal agressor. Estas foram analisadas, por meio dos testes Qui-Quadrado de independência e Exato de Fisher, com nível de significância de 5%. Obteve-se, também, intervalos de confiança para as variáveis em estudo, considerando um nível de confiança de 95% no *software* R versão 3.1.1 (R CORE TEAM, 2014), e os resultados dispostos em tabelas e gráficos.

## 4 Resultados e Discussão

A avaliação do perfil dos pacientes de atendimento anti-rábicos procedeu-se com uma análise descritiva e inferencial.

Por meio dos gráficos de densidade, Figura 2 (a), observa-se que a idade dos pacientes de atendimento anti-rábicos, com maior frequência de ocorrência é na idade de 0 a 10 anos para o gênero masculino e adulta para as mulheres. Essa inversão ocorre, pois na infância os meninos tem um maior contato com cães e gatos, enquanto na idade adulta, as mulheres são responsáveis pelos trabalhos domésticos, o que inclui cuidado com os mesmos. Em relação a população idosa, a proporção de atendimentos é baixa, assim como sua presença na população, entretanto, na infância e idade adulta, essa proporcionalidade não se verifica. São mais prováveis os atendimentos na infância do que na idade adulta, uma vez que na infância a proporção de atendimentos é maior do que a esperada, e o contrário ocorre na idade adulta.

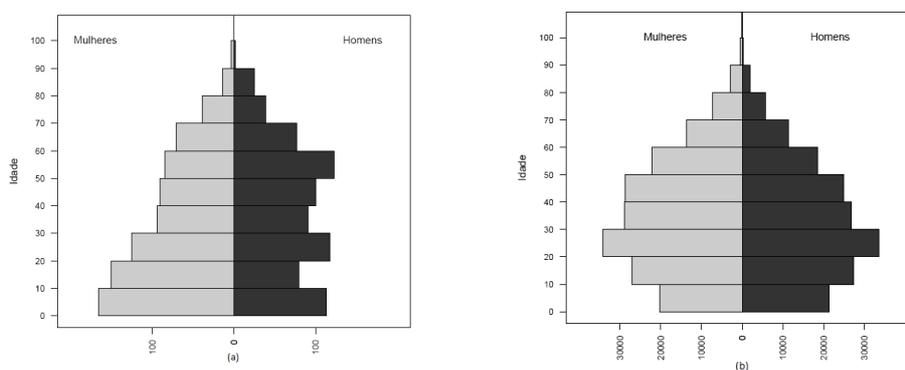


Figura 2: Representação gráfica da densidade da frequência dos pacientes atendidos (a) e da população de Maringá (b).

Observando a amostra de pacientes, nota-se que a proporção de indivíduos atendidos do gênero masculino é superior ao feminino, (valor-p = 0,0107). Considerando as pacientes, dentre a amostra, verificou-se que 2,3% eram gestantes. Os cuidados com essas pacientes serão os mesmos que para os demais, uma vez que as gestantes no início da gravidez realizam todos procedimentos preventivos necessários.

A maior proporção dos pacientes atendidos são da cor branca (74,4%), sendo este um resultado esperado, pois de acordo com o CENSO 2010 a população da cidade de Maringá-PR, é predominantemente branca. Em relação ao grau de escolaridade, observou-se que o ensino médio completo é o mais frequente (14,5%). A maioria dos pacientes residem na zona urbana de Maringá (95,72%), confirmando os resultados encontrados por Carvalho, Soares e Franceschi (1997). Para verificar qual a verdadeira proporção dos pacientes que residem na zona urbana, calculou-se um intervalo de confiança. Os resultados evidenciaram a proporção de pacientes

residentes na zona urbana está entre 94,62% e 96,65%.

O tipo de exposição com maior proporção de ocorrência é a mordedura, (entre 77,86% e 81,80%). E o tipo de animal que provocou a maior parte dos casos foi o cão (91,56 %) e que esta proporção está entre 90,18% e 92,92%.

Os locais com maior frequência de ferimentos são as mãos e pés (39,79%). Esta ocorrência se deve ao fato da maioria dos ferimentos serem provocados cães e gatos, corroborando com os resultados encontrados por Carvalho, Soares e Franceschi (1997). O tipo de ferimento mais comum nos pacientes atendidos são os superficiais, como por exemplo os arranhões ocasionados por unhas ou dentes de cães ou gatos, uma vez que esses são os animais mais frequentes, conforme visto anteriormente. Quanto ao procedimento adotado com maior frequência para o tratamento de um indivíduo atacado por um animal é a observação e vacina (57,7%). Em que a observação consiste em deixar o animal sob vigilância por 10 dias após a exposição, verificando o estado do mesmo. Assim, se a suspeita de raiva for descartada deve-se suspender o esquema profilático e encerrar o caso, caso-contrário deve-se adotar as medidas adequadas para o tratamento. Verificou-se ainda que os casos de atendimentos anti-rábicos registrados ocorreram com maior frequência no segundo semestre de 2012.

Por meio da Tabela 1 verificou-se que o tipo e a localização do ferimento apresentam uma associação (valor-p < 0,0001). Sendo que os ferimentos profundo e superficial ocorrem frequentemente nas mãos/pés, sendo estes os locais mais atingidos por cães e gatos, conforme apresentado anteriormente.

Tabela 1: Frequência da relação do tipo de ferimento quanto à localização do ferimento.

Ferimento	Localização do ferimento					
	Mucosa	Cabeça/pescoço	Mãos/pés	Tronco	Membros Inf.	Membros Sup.
Profundo	15	52	306	19	90	162
Superficial	14	64	306	47	152	305
Dilacerante	4	6	33	3	9	19

Os cães provocam, a maioria dos ataques, com ferimentos superficiais. Porém, ao se realizar o teste qui-quadrado de independência verificou-se que não há associação entre o tipo de animal e ferimento (valor-p = 0,7432). Na Tabela 2, verifica-se que há associação entre o tipo de animal e o número de ferimentos (valor-p < 0,0001), destacando-se os cães que provocam ferimentos únicos.

Tabela 2: Frequência da relação do tipo de animal quanto ao número de ferimentos.

Ferimento	Tipo de animal				Total
	Canina	Felina	Quiróptera	Primata	
Único	880	41	9	6	936
Múltiplo	602	54	0	6	662
Sem ferimento	4	1	3	1	9
Total	1486	96	12	13	1607

## 5 Conclusões

O perfil dos pacientes atendidos são predominantemente do sexo masculino, cor branca, possuem ensino médio completo e residem na zona urbana de Maringá. Existe relação entre a localização e o tipo de ferimento e tipo de animal e o número de ferimentos. Ainda, por meio das análises, verificou-se que as crianças (0 a 10 anos) mais atingidas são do gênero masculino. Na idade adulta, essa situação se inverte, sendo mulheres as mais atingidas. Os acidentes mais comuns são as mordeduras de cães, com ferimentos únicos e superficiais, sendo as mãos e pés as partes do corpo mais atingidas.

## Referências

- [1] BABBONI, S. D.; MODOLO, J. R. Raiva: origem, importância e aspectos históricos. *UNOPAR Cient Ciênc Biol Saúde* 2011;13(Esp), p. 349-56, 2011.
- [2] CARVALHO, W. O.; SOARES, D. F. P. P.; FRANCESCHI, V. C. S. Características do atendimento prestado pelo serviço de profilaxia de Raiva Humana na rede municipal de saúde de Maringá-Paraná, no ano de 1997. *Informe Epid. do SUS*. v.11, p.25-35, 2002.
- [3] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. 2013. Acesso em 20 julho. 2014.
- [4] WADA, M. Y.; ROCHA, S. M.; MAIA-ELKHOURY, A. N. S. Situação da Raiva no Brasil, 2000 a 2009. *Epidemiol. Serv. Saúde, Brasília*, v. 20, n. 4, dez., 2011.

# Tratamento de Dados Faltantes: Uma avaliação de métodos de imputação de dados a partir do desempenho de um classificador Bayesiano

Laércio Ives Santos<sup>1</sup>

Maria de Fátima Ferreira Almeida<sup>2</sup>

**Resumo:** O presente trabalho aborda um estudo comparativo do comportamento de métodos de imputação de dados aplicados em dois conjuntos de dados do repositório UCI tomando como base o desempenho de um classificador Bayesiano. Os resultados mostram que para um dos conjuntos de dados estudados o desempenho do classificador melhora quando os métodos de imputação são usados.

**Palavras-chave:** *Missings Values*, Imputação de Dados, Classificação Bayesiana.

## 1 Introdução

Nos últimos anos muitos métodos para tratamento do problema da falta de dados em variáveis(*missings values*) têm sido estudados e metodologias foram desenvolvidas. Esse problema pode ser causado por diversos motivos, dentre eles podem ser citados problemas com o armazenamento dos dados e o não preenchimento cadastral. Uma estratégia que muitos pesquisadores usam para amenizar o problema é o descarte da amostra. O que não é viável em alguns modelos estatísticos, tendo em vista o quantitativo de *missings*.

Algo muito utilizado para amenizar a falta de dados é a substituição dos mesmo por dados inferidos por algum mecanismo. Esse processo é denominado Imputação.

Vários métodos podem ser usados para realizar a tarefa de Imputação. Nesse trabalho optou-se por usar 4 desses. Que são: PMM(*Predictive mean matching*), NORM(*Bayesian linear regression*) e NORM.nob(*Linear regression, non-Bayesian*), Mean.

No método PMM os parâmetros são estimados a partir de uma distribuição a posteriori própria. São calculados os valores preditos para os dados observados e *missings values*. Para cada *missing* predito, procura-se a unidade observada com valor predito mais próximo, e utiliza-se o valor observado como valor a ser imputado.

Os métodos NORM e NORM.nob trabalham com Regressão Linear e consiste em predizer o valor de uma determinada variável(predita) em relação ao valor de outra variável(preditora). Sendo que nesse caso a variável predita sempre é dependente da variável preditora. No caso da regressão Linear Bayesiana a variável que não foi amostrada é predita tomando como base os valores das

---

<sup>1</sup>IFNMG: Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais, Montes Claros-MG. Email: [laercio.santos@ifnmg.edu.br](mailto:laercio.santos@ifnmg.edu.br)

<sup>2</sup> IFNMG: Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais, Montes Claros-MG. Email: [fatima.almeida@ifnmg.edu.br](mailto:fatima.almeida@ifnmg.edu.br)

demais variáveis da amostra usando uma abordagem baseada na teoria da decisão de Bayes.

O método *Mean* consiste basicamente em prever os valores da variável faltante tomando como valor a média aritmética das demais variáveis da amostra.

A tarefa de Classificação é uma técnica computacional utilizada que pode criar modelos que descrevem classes de dados. Esses modelos são baseados em uma análise de um conjunto de dados de treinamento, onde as classes dos dados são conhecidas [WITTEN, 2003 apud FONSECA, 2007]. Após a geração do modelo, o mesmo pode ser aplicado em um conjunto de dados de teste, onde as classes de dados são previstas pelo modelo e avaliadas em relação ao valor da classe real.

Um modelo de classificação de dados é desenvolvido em duas etapas. A primeira etapa é denominada etapa de treinamento, em que, uma amostra de dados, denominada amostra de treinamento, é submetida ao classificador. Cada registro da amostra pertence a uma determinada classe, determinada por um de seus atributos, esse atributo é denominado atributo classe. A classificação é um tipo de aprendizado supervisionado.

A segunda etapa do modelo é a etapa de testes, nela, um subconjunto da amostra é usado para validar o modelo criado na etapa de treinamento.

Um modelo de classificação Bayesiana é uma abordagem estatística que pode prever a probabilidade de um registro pertencer a uma determinada classe. Ele oferece uma poderosa técnica de classificação supervisionada assumindo que todos os atributos de entrada possuem a mesma importância e são independentes entre si.

## 2 Objetivos

O objetivo desse trabalho é comparar o comportamento de um classificador baseado na teoria de decisão de Bayes ao trabalhar com amostras de dados incompletas e completadas por algum método de imputação dos dados faltantes.

## 3 Metodologia

Esse estudo aborda uma avaliação quantitativa de alguns métodos de imputação de dados em amostras usadas para o treinamento de um classificador Bayesiano.

Para tanto, o estudo foi realizado em quatro etapas:

1. Implementação do classificador Bayesiano na linguagem R.
2. Teste da eficiência do classificador em 2 conjuntos de dados, medindo as taxas de erro em cada um.
3. Teste da eficiência do classificador na presença de ruído (*missings values*) nos 2 conjuntos de dados;
4. Teste do classificador nos 2 conjuntos completados com dados inferidos por métodos de imputação;

## 4 Resultados e discussão

O classificador foi implementado utilizando a linguagem R[R, 2013], devido a facilidade de manipulação de funções estatísticas que a linguagem oferece. Em seguida, foram realizados vários experimentos em duas bases de dados Iris e pima-indians-Diabetes ambas do repositório UCI[FRANK & ASUNCION, 2013]. O objetivo é analisar a eficiência do classificador quando aplicado nas bases de dados completas, faltando dados, e completadas a partir de do uso dos métodos de Imputação.

No primeiro experimento, o classificador foi testado na base de dados Iris, utilizando apenas duas das três classes que compõe a amostra. A base de dados Iris possui cinco atributos, sendo quatro contínuos e um categórico (atributo que define a classe de cada linha). Para duas classes a amostra contém 100 linhas, dessas, foram utilizadas 80 para treinamento e 20 para testes. Ao ser testado com a base de dados completa, o classificador obteve uma eficiência de 100% de acertos.

No segundo experimento, a base de dados Iris foi submetida a testes com sendo introduzido um percentual de ruído(*missings values*). O resultado é mostrado no gráfico da Figura 1. Nesse gráfico é mostrada a evolução do percentual de erro do classificador quando introduzido um ruído de 1 a 60%. Pode ser notado que o classificador começa a perder a eficiência em taxas de ruído maiores que 20%.

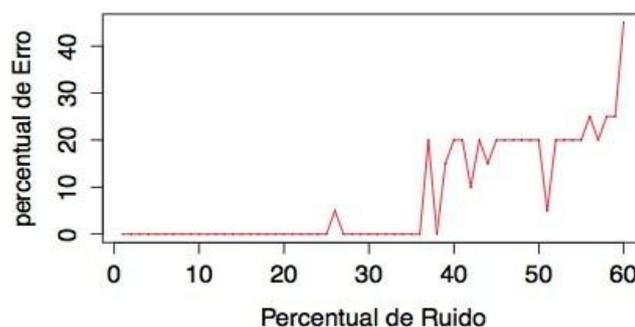


Figura 1: Evolução da taxa de erro em relação a quantidade de missings values

O terceiro experimento realizado na base de dados Iris, objetivou avaliar 3 métodos de tratamento de missings values: PMM, NORM, NORM.nob, utilizando taxas de ruído de 5, 10 e 20%. Em todos os casos a taxa de acertos do classificador se manteve em 100%. O que pode ser atribuído ao fato de a base ser normal em dois atributos e ter um valor baixo para o desvio padrão em todos os atributos.

No quarto experimento o classificador foi testado na base de dados pima -indians-Diabetes, utilizando 500 linhas da base de dados para treinamento e 268 para testes. Nesse experimento, o classificador obteve uma taxa de acertos de 79,36%. Levando em consideração que nenhum de seus atributos passou no teste de normalidade e ainda que a base de dados está incompleta.

No quinto e último experimento, a base de dados pima-indians-Diabetes foi completada usando os métodos de imputação: PMM, NORM, NORM.nob e Mean os resultados são mostrados na Tabela 1. A notação adotada foi, Método-N, em que N é a quantidade de iteração no processo de

imputação.

Existe um pequeno aumento na eficiência de um maneira geral do Classificador quando se usam os métodos de imputação para completar o conjunto de dados. E ainda é importante notar que, quando a quantidade de iteração dos métodos Norm e Norm.nob foi reduzida, há uma melhora no desempenho do Classificador.

Tabela 1: Avaliação Quantitativa dos métodos de imputação do pacote MICE do *Software R* associado ao classificador Bayesiano

Método de imputação	Taxa de erro do classificador Bayesiano (%)
Faltando dados	21.64
PMM – 5	20.15
Norm – 5	20.52
Mean – 5	21.27
norm.nob – 5	20.52
PMM – 1	20.89
Norm – 1	19.03
Mean – 1	21.26
norm.nob – 1	19.03

## 5 Conclusão

Após efetuados os testes do classificador desenvolvido nesse trabalho nas duas base de dados, pode-se notar uma menor taxa de erro na amostra Iris. Que dentre as amostras utilizadas é a única que continha variáveis normais. Em contra partida na base de dados pima-Indian-diabetes, que todas as variáveis não são normais o classificador obteve uma taxa de erros relativamente alta.

O desempenho do Classificador melhorou na base de dados pima-indian-Diabetes quando utilizados métodos de imputação de dados. Importante salientar que o princípio dos métodos norm e norm.nob assim como o do classificador é a teoria da decisão Bayesiana. O que pode ter influenciado na melhora do desempenho do classificador quando utilizados esses dois métodos.

Dessa forma, observamos que o desempenho do classificador na base de dados Iris é pouco influenciado pela presença ou não de *missings*, visto que somente a partir de 20% de ruído que tal desempenho começa a cair. Entretanto ao usar métodos de imputação na base de dados pima-indian-Diabetes existe uma leve melhora no desempenho do classificador. Principalmente quando usados os métodos baseados na teoria de decisão de Bayes, Norm e Norm.nob. Assim, concluímos que para essa base de dados ambos os métodos são mais eficientes que os demais estudados.

## Referências

FONSECA, M. P. A. *Classificação Bayesiana de grandes massas de dados em ambientes ROLAP*. Tese de doutorado apresentada ao programa de pós-graduação de engenharia da universidade federal do rio de janeiro rio de janeiro .2007.

FRANK, A. ;ASUNCION, A. *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml>].  
Irvine, CA: *University of California, School of Information and Computer Science*. 2012.

R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. R  
*Foundation for Statistical Computing, Vienna, Austria*. ISBN 3-900051-07- 0, URL  
<http://www.R-project.org/>. 2013.

# Impactos das fusões e aquisições na rentabilidade das empresas adquirentes: uma análise do setor de tecnologia da informação brasileiro no período de 1994 a 2013

Camila Batista Nascimento<sup>1 5</sup>

Gabriel Rodrigo Gomes Pessanha<sup>2 5</sup>

Luciene Resende Gonçalves<sup>3 5</sup>

Jennifer Elis Pires<sup>4 5</sup>

**Resumo:** Esse trabalho tem por objetivo identificar a existência de impactos das estratégias de Fusões e Aquisições (F&A) na rentabilidade das empresas adquirentes do setor de Tecnologia da Informação. O crescente volume das transações de Fusões e Aquisições envolvendo empresas brasileiras demonstra o papel que essa estratégia vem assumindo em questões de aumento de competitividade e geração de valor aos acionistas, deste modo se torna de total importância a pesquisa desenvolvida. A pesquisa está compreendida entre os anos de 1994 a 2013. Realizou-se a metodologia de análise de intervenção nas variáveis de rentabilidade das empresas envolvidas nos processos de F&A, com o objetivo de detectar os impactos dessas operações na rentabilidade das empresas adquirentes. Os resultados encontrados demonstram que de acordo com a característica intervencionista, não são todos os eventos de F&A que conseguem repercutir na rentabilidade das empresas do setor de TI, contudo, quando um evento modifica a rentabilidade da organização em sua maior parte é de forma positiva.

**Palavras-chave:** *intervenção, rentabilidade, fusão.*

## 1. Introdução

Os processos de Fusões e Aquisições (F&A) no Brasil, e no resto do mundo, apresentaram nas últimas décadas grandes transformações em seu volume, demonstrando crescente aumento. Essa mudança segundo especialistas da área, é fruto da modificação que vem ocorrendo na economia mundial, em um ambiente globalizado, onde as organizações vêm buscando, cada vez mais, competitividade.

Há muitos motivos para a realização das estratégias de F&A para as empresas. Entre eles estão a necessidade de crescimento, o aumento de *Market share*, a criação de barreiras à entrada de novos concorrentes, aquisição de tecnologia, realização de sinergias e obtenção de economias de escala. Pessanha et al. (2012) acredita, contudo, que independente dos motivos que levam as empresas a adotarem como estratégias F&A o objetivo principal de tais eventos é gerar e ampliar valor às

---

<sup>1</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha-MG. Email: [camila\\_batista@hotmail.com](mailto:camila_batista@hotmail.com)

<sup>2</sup> ICSA – UNIFAL: Universidade Federal de Alfenas Varginha-MG. Email: [gabriel.pessanha@unifal-mg.edu.br](mailto:gabriel.pessanha@unifal-mg.edu.br)

<sup>3</sup> ICSA – UNIFAL: Universidade Federal de Alfenas-MG. Email: [luciene.goncalves@unifal\\_mg.edu.br](mailto:luciene.goncalves@unifal_mg.edu.br)

<sup>4</sup> ICSA – UNIFAL: Universidade Federal de Alfenas-MG. Email: [jenni.elis@hotmail.com](mailto:jenni.elis@hotmail.com)

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

empresas e aos acionistas, ressaltando que há dificuldade em determinar e mensurar a criação de valor para a organização e, principalmente, para os acionistas.

Nos últimos anos o setor de Tecnologia da informação (TI) é o que mais registra processos de F&A no Brasil, segundo dados coletados pela consultoria KPMG (2013). Posto isto, torna-se proeminente verificar os impactos sucedidos em empresas de TI que passaram por esses processos.

As empresas de TI começaram a se destacar nos processos de F&A durante a crise *subprime* que aconteceu nos Estados Unidos em 2008. Silva (2011) diz que ainda que não existam provas que essa crise tenha sido motivo para grandes transformações nas estruturas das empresas de TI, há relatos de que depois da crise o governo Brasileiro disponibilizou auxílio, pelo BNDES (Banco Nacional do Desenvolvimento), o que possibilitou operações de F&A no setor através concessão de crédito subsidiário.

## **2. Objetivo**

Este trabalho tem como objetivo verificar se as estratégias de fusões e aquisições realizadas por empresas no setor de TI no Brasil nos últimos anos tem impacto na rentabilidade das mesmas. Com o intuito de auxiliar o esclarecimento das hipóteses formuladas relata-se a metodologia aplicada para a verificação do problema e em seguida os resultados dos métodos empregados e por fim as considerações finais.

## **3. Metodologia**

Com a finalidade de identificar os efeitos dos processos de fusão e aquisição sobre a rentabilidade das empresas de tecnologia da informação adquirentes, foi aplicada a análise de intervenção. Para a utilização da metodologia de Box e Jenkins (1976), conhecida como modelos Arima (autorregressivos e de médias móveis), e análise de intervenção. Os dados para análise do estudo foram retirados do banco de dados do Economática, e foram utilizadas séries trimestrais contidas entre os anos de 1994 e 2013, para as variáveis ROA (retorno dos ativos), ROE (retorno do patrimônio líquido), ML (margem líquida) e séries diárias contidas no mesmo período de tempo para a variável retorno das ações. Para a análise dos dados coletados foram utilizados os *softwares Statistica e Gretl*.

Para as variáveis ROA, ROE e ML, somente as empresas: Ideiasnet, Itaotec e Totvs, apresentaram séries possíveis de captar as intervenções, as séries dessas variáveis das demais empresas não possibilitaram o ajuste. E para a variável retorno das ações, as empresas possíveis de ajustes foram: Belmatech, Ideiasnet, Itaotec, Positivo Informática e Senior Solution. As intervenções foram definidas de acordo com a ocorrência da fusão ou da aquisição nas empresas do setor de TI. Foram analisadas 60 operações de fusões e aquisições.

## **4 Análise e discussão**

O primeiro passo para a análise foi verificar visualmente através do gráfico da série original e as funções de autocorrelação (FAC) e autocorrelação parcial (FACP), se a série possuía visualmente a presença de algum componente. Após a análise visual foi aplicado o teste de raiz unitária, Dickey-

Fuller Aumentado. Comprovando a estacionariedade das séries. Após a constatação, foram gerados os modelos ajustados das séries estudadas, e escolhidos os melhores modelos através do Critério de Informação de Akaike e o Critério do Erro Quadrático Médio de Previsão (EQMP), demonstrados nas Tabelas de 1, 2, 3 e 4. Para verificar se o modelo ajustado constitui um ruído branco, foi utilizado o teste de Ljung-Box. Assim verificando que o modelo mais adequado é estacionário.

Após a confirmação de que as séries são estacionárias, de forma geral, foi possível verificar a existência de correlação significativa tanto no primeiro *lag* das funções de autocorrelação como no primeiro *lag* das funções de autocorrelação parcial, deste modo propondo um modelo com estrutura autorregressiva e de médias móveis de ordem 1. A partir da análise do desempenho gráfico da série, reconheceu-se, que a intervenção ocorreu de forma abrupta e temporária. O que levou a concluir que a variável *dummy* assumiu valor 1 no momento em que a F&A ocorreu e assumiu valor zero quando não houve fusões e aquisições.

$$X_{i,t} \begin{cases} 1, \rightarrow T = \text{Ocorrência de fusões e aquisições} \\ 0, \rightarrow T = \text{Não ocorrência de fusões e aquisições} \end{cases}$$

Tabela 1: Estimativas dos parâmetros, erros padrões e valor P de modelos arima com intervenções para a variável retorno do ativo (ROA) para as empresas: Ideiasnet, Itautec e Totvs.

Modelo	Estimativas		Intervenções				
Ideiasnet Arima (1,0,1)	p(1)	q(1)	(29) Om(1)	Del(1)	(30) Om(2)	Del(2)	
	<b>,9360</b>	<b>,6873</b>	<b>,0401</b>	<b>,6410</b>	<b>,0129</b>	<b>,6410</b>	
	(,0723)	(,1425)	(,1995)		(,2200)	2,7011)	
Erro Padrão			(1,8380)				
Valor P	0,0000	0,0000	0,8413		0,9534	0,8130	
			0,7280				
Itautec Arima (0,0,2)	q(1)	q(2)	(45) Om(1)	Del(1)	(49) Om(2)	Del(2)	(51) Om(3)
	<b>-,4507</b>	<b>-,2338</b>	<b>,0392</b>	<b>,9076</b>	<b>,0219</b>	<b>,9353</b>	<b>,0233</b>
	(,1408)	(,1123)	(,0366)		(,0472)		<b>,8747</b>
Erro Padrão		(,2997)		(,2230)		(,0443)	(,4420)
Valor P	0,0021	0,0411	0,2880		0,6434		0,6025
			0,0034		0,0000		0,0523
Totvs Arima(0,0,1)	q(1)		(13) Om(1)	Del(1)	(16) Om(2)	Del(2)	(19) Om(3)
	<b>,3770</b>		<b>,0618</b>	<b>,8505</b>	<b>-,0018</b>	<b>-,1830</b>	<b>,0239</b>
	(,1753)		(,0578)	(,3796)	(,0088)	(,3793)	(,0736)
Erro Padrão		(12,1390)					
Valor P	0,0433		0,2967	0,0348	0,8406	0,0053	0,7478
							0,9846
Erro Padrão			(20) Om(4)	Del(4)	(21) Om(5)	Del(5)	(22) Om(6)
			<b>,0530</b>	<b>,5737</b>	<b>,0465</b>	<b>1,0539</b>	<b>-,0502</b>
			(,2971)	(2,1926)	(,0905)	(,1825)	(,0858)
Valor P		(1,146)					
			0,8588	0,8588	0,6072	0,0000	0,5663
			0,7548				

As estimativas dos parâmetros autorregressivos (AR), de médias móveis (MA) e da variável de intervenção são significantes para algumas empresas e para outras não, conforme resultado das Tabelas (1, 2, 3 e 4). O teste de Ljung e Box aceita a hipótese nula de ruído branco, ou seja, da não

presença de autocorrelação significativa nos resíduos. O que demonstra que os modelos ajustados são ruído branco.

Ao analisar a variável retorno sobre o ativo (ROA) para a empresa Ideiasnet, pode-se afirmar, por meio da Tabela 1, que apenas os parâmetros p(1) e q(1) tiveram significância estatística, apresentando os valores P de 0, os parâmetros das intervenções abruptas temporárias nas observações 29 (Om (1) e Del (1)) e 30 (Om(2) e Del(2)) não apresentaram significância estatística. A observação 29 corresponde à F&A das empresas Carvopar Participações, Bolsa de Mulher e Virtualab Holding, e a observação 30 a operação de fusão e aquisição da Officer, pela empresa Ideiasnet.

A partir da Tabela 1, pode-se observar que em nenhuma série as operações demonstraram uma alteração no comportamento da série original a ponto de se converterem em uma intervenção. Este fato poderia ser justificado pelo tamanho das empresas adquiridas, ressaltando que na maioria dos casos as empresas adquiridas do setor de TI são de pequeno porte, sendo este um possível motivo para a insignificância estatística das operações de F&A nas empresas do setor de tecnologia da informação.

Tabela 2: Estimativas dos parâmetros, erros padrões e valor P de modelos arima com intervenções para a variável retorno do patrimônio líquido (ROE) para as empresas: Ideiasnet, Itautec e Totvs.

Modelo	Estimativas		Intervenções				
			(29) Om(1)	Del(1)	(30) Om(2) Del(2)	(30) Om(2) Del(2)	(51) Om(3) Del(3)
Ideiasnet Arima (1,0,1)	p(1) .9408	q(1) .6958	.0878	.6456	.0395	.6456	
Erro Padrão	(.0716)		(.2420)		(.2630)		
Valor P	(.1484)		(.9387)		(1,148)		
	0,0000		0,7184	0,4950	0,8811		
	0,0000				0,5765		
Itautec Arima (1,0,0)	p(1) .4482		Om(1) Del(1) .0934	.9146	Om(2) Del(2) .0510	.9326	Om(3) Del(3) .0424 .8917
Erro Padrão	(.1100)		(.1018)		(.1280)		(.1252) (.4607)
Valor P			(.2741)		(.2545)		
	0,0001		0,3623	0,0013	0,6916		0,7357 0,0570
					0,0004		
Totvs Arima (0,0,1)	q(1) -.3221		Om (1) Del (1) .1515	.9842			
Erro Padrão	(.1389)		(.0649)	(.0584)			
Valor P							
	0,0271		0,0263	0,0000			

Outro possível motivo para essa insignificância é o fato da variável ROA ser baseada no ativo total da empresa, conforme Pessanha et al. (2012) esta variável examina a competência da administração na criação de lucros com seus próprios ativos disponíveis, deste modo pode-se concluir que o evento de fusão e aquisição prejudicou a geração de lucros. O que pode justificar a não captura de uma intervenção no momento da fusão ou aquisição.

Ao se analisar a variável ROE (Tabela 2), somente a empresa Totvs, apresentou uma operação de F&A significativa estatisticamente, o parâmetro q(1) apresentou o valor P de 0,0271, e os parâmetros da intervenção abrupta temporária na observação 21 (Om (1) e Del (1)) apresentaram

significância estatística, com valor P de 0,0263 e 0,0000 respectivamente. A observação 21 corresponde à operação de F&A da Totvs com as empresas *Henry Software* e 70% do Totalbanco.

Em relação a variável ML a empresa Itautec apresentou o evento correspondente a observação 51 significativo, a observação 51 corresponde à operação de F&A da empresa Itautec com a Itautec Informática. Dirigindo a análise para a variável retorno das ações, apenas as empresas Belmatech e Itautec apresentaram eventos significativos. A observação 177 representa o evento de F&A da empresa Belmatech com a Rentec. A observação 2298 equivale a F&A que ocorreu entre a empresa Ideiasnet e a empresa Ciashop. Após a análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) depois do ajuste para cada uma das séries estudadas, somada a confirmação da hipótese nula de ruído branco pelo teste de LB, demonstraram que as dinâmicas das séries estudadas estão sendo bem explicadas pelos coeficientes dos modelos ajustados.

## 5. Considerações finais

Os impactos que as F&A representaram nas empresas de TI foram avaliados pela metodologia de Box e Jenkins (1976) com a análise de intervenção, a partir desta metodologia buscou encontrar as F&A que tenham levado a uma alteração no comportamento da série original a ponto de se tornarem uma intervenção. Os resultados encontrados demonstraram que apenas três empresas se mostraram sensíveis estatisticamente às fusões e aquisições, e entre elas somente a empresa Belmatch em uma das operações de F&A, envolvendo as empresas Snackcontrol Sistema, Logic Controls e MisterChef apresentou uma intervenção negativa em sua rentabilidade das ações, esse resultado demonstra o caráter positivo das intervenções para a rentabilidade das empresas de TI após os eventos de F&A.

Cabe aqui ressaltar as limitações deste trabalho, em que são analisadas apenas operações de curto prazo, deste modo eventos de F&A que poderiam ter causado uma mudança na rentabilidade das empresas de TI em longo prazo, não foram captadas como intervenção no curto prazo. Somado ao caráter probabilístico da metodologia aplicada, o que pode ter proporcionado um viés aos resultados.

## Referências

- [1] BOX, G. E. P.; JENKINS, G. M. *Time series analysis: forecasting and control*. San Francisco: Holden-Day, 1976.
- [3] MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. 2. Ed. São Paulo: Edgard Blucher, 2006.
- [4] PESSANHA, G. R. G.; CALEGARIO, C. L. L.; SÁFADI, T.; AZARA, L. N. Impactos das estratégias de fusão e aquisição na rentabilidade dos bancos adquirentes: uma aplicação dos modelos de intervenção no setor bancário brasileiro. *Revista de Administração Mackenzie*. São Paulo, v. 13, n° 5, set/out. 2012.
- [5] SILVA, J. *Análise da variação de valor das empresas brasileiras de tecnologia de informação e comunicação durante a crise do subprime*. 2011. Trabalho de Conclusão de Curso (Graduação em Administração) – Escola de Administração, UFRGS, Porto Alegre, 2011.

## Estudo probabilístico das condições de fertilidade do solo: Uma análise geoestatística sob a perspectiva de associação de Krigagem Indicativa e álgebras de mapas

Maria de Fátima Ferreira Almeida– IFNMG<sup>1 6</sup>

Wagner Rogério Ferreira Pinheiro<sup>2 6</sup>

Magaly Stefânia Almeida - FACIT, MG<sup>3 6</sup>

Gérson Rodrigues dos Santos – DET, UFV<sup>4 6</sup>

Domingos Sávio Valente – DEA, UFV<sup>5 6</sup>

**Resumo:** A Geoestatística, através da técnica de Krigagem Indicativa é de grande importância na agricultura de precisão para fazer planejamento de correção do solo de forma localizada e identificar zonas de manejo. Neste trabalho utilizou a Krigagem Indicativa no planejamento de manejo de correção localizada nos pontos que apresentaram carência dos nutrientes P, Ca, Mg, K e pH fora dos níveis de suficiências adequados ao cultivo obtido por meio de álgebra de mapas associados aos mapas de Krigagem Indicativa, numa área já cultivada com o café da Zona da Mata Mineira.

**Palavras-chave:** Krigagem Indicativa, propriedades químicas, solos , mapas.

### 1. Introdução

De acordo com Machado et al. (2007) na agricultura brasileira, atualmente, grandes áreas são consideradas homogêneas; nelas a necessidade média de utilização dos insumos promove o uso de doses iguais de fertilizantes, desconsiderando a variabilidade espaço temporal, podendo sobrecarregar uma gleba que é fértil e não atingir níveis ótimos para uma alta produtividade em outras glebas deficientes.

Neste trabalho utilizou a técnica de Krigagem Indicativa no planejamento de manejo de correção de solo localizada nas subáreas que apresentaram os nutrientes: fósforo (P), cálcio (Ca), magnésio (Mg), potássio (K) e pH fora dos níveis de suficiência adequados ao cultivo de café. Esta

---

<sup>1</sup>DET – UFV: Estudante do curso de Pós - Graduação em Estatística aplicada da Universidade Federal de Viçosa -MG.  
Email: [fairam2006@yahoo.com.br](mailto:fairam2006@yahoo.com.br)

<sup>2</sup>DET – UFV: Estudante do curso de Pós - Graduação em Estatística aplicada da Universidade Federal de Viçosa -MG.  
Email: [wagner22356@gmail.com](mailto:wagner22356@gmail.com)

<sup>3</sup>FACIT: Estudante da Faculdade de Ciências e Tecnologia de Montes Claros. Email: [magalystefania@yahoo.com.br](mailto:magalystefania@yahoo.com.br)

<sup>4</sup>DET – UFV: Professor do Programa de Pós Graduação da Universidade Federal de Viçosa -MG.  
Email: [geron.santos@ufv.br](mailto:geron.santos@ufv.br)

<sup>5</sup>DEA – UFV: Professor do Departamento de Engenharia Agrícola da Universidade Federal de Viçosa

análise foi feita para uma área já cultivada com o café no Município de Araponga na região da Zona da Mata Mineira.

## 2. Material e métodos

### 2.1. O estimador de Krigagem Indicativa para atributos numéricos

De acordo com Landim (2003) utilizando o método de Krigagem Indicativa é possível elaborar estimativas de proporções para determinados valores que estão abaixo ou acima de um determinado nível de corte (percentis) de uma distribuição acumulada em determinado local; e, em seguida transformá-los em indicadores definidos pela expressão(1) obtido a partir do semivariograma experimental (2) das amostras:

$$I_i(z_c) = \begin{cases} 1, & \text{se } z(x) \leq z_c \\ 0, & \text{se } z(x) > z_c \end{cases} \quad (1)$$

$$\hat{\gamma}(\mathbf{h}; z_c) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [I(\mathbf{x} + \mathbf{h}; z_c) - I(\mathbf{x}; z_c)]^2 \quad (2)$$

onde:  $z(\mathbf{x})$  = valor observado e  $z_c$  = valor do nível de corte e

a função de distribuição acumulada dos valores abaixo do nível de corte é definida por:

$$\hat{F}(z_c) = \frac{1}{n} \sum_{i=1}^n \lambda_i I_i(z_c) \quad (3)$$

Em que:  $\lambda_i$  são os pesos,  $I_i$  os indicadores,  $n$  o número de pontos e  $z_c$  o nível de corte. Em que

$\sum_{i=1}^n \lambda_i = 1$  é a condição de não viés.

A partir da aplicação da formulação do semivariograma Indicativo retorna os valores estimados de  $\gamma(h)$  de 0 a 1, em função da distância ( $\mathbf{h}$ ), para  $\mathbf{h} \leq \mathbf{a}$  (alcance).

### 2.2. Descrição da variável

Os dados do solo foram medidos em uma fazenda localizada em Araponga, estado de Minas Gerais, Brasil (20 ° 42'33 "S, 42 ° 34'17" W, altitude média de 913 m de altitude). A fazenda tem 86 hectares plantados com Coffea arabica L. A fazenda está localizada em uma área montanhosa, e o solo é classificado como Latossolo Vermelho distrófico (Staff Solos,2006 citado por Valente et al., 2012).

Os pontos de cortes foram estabelecidos considerando o ponto médio do extremo inferior dos níveis de suficiência das variáveis K, Mg, Ca, P e pH propostos por Alves (2012).

## 3. Resultados e discussão

A Tabela 1 mostra os valores dos parâmetros dos ajustados dos semivariogramas referentes a cada variável estudada.

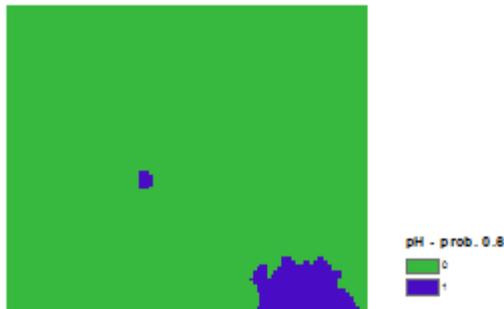
Tabela 1 - Parâmetros dos semivariogramas binários para os nutrientes: Acidez/alcalinidade (pH), Fósforo (P), Magnésio (Mg), Potássio (K) e Cálcio (Ca) no cultivo de café.

Nutrientes	Modelo	Efeito pepita (C0)	Patamar (C)	Alcance (a)	IDE (Zimback)	Erro Padrão (RSS)
$K \geq 65 \text{mg/dm}^3$	Exp.	0.09693	0.179076	60,3	0,9351	$3,500 \times 10^{-3}$
$K \leq 73 \text{mg/dm}^3$	Exp.	0.117897	0.237811	227,66	0,504	$4,115 \times 10^{-3}$
$Ca \geq 1.9 \text{cmol}_c/\text{dm}^3$	Exp.	0,112478	0,174187	156,058	0,373	$3,875 \times 10^{-3}$
$Ca \leq 2.25 \text{cmol}_c/\text{dm}^3$	Exp.	0,09539	0,23378	37,5	0,592	$1,676 \times 10^{-3}$
$Mg \geq 0,9 \text{mg/dm}^3$	Exp.	0,144168	0,1842	90,55	0,218	$8,121 \times 10^{-3}$
$Mg \leq 1,0 \text{mg/dm}^3$	Exp.	0,142476	0,23386	85,14	0,403	$5,435 \times 10^{-3}$
$P \geq 15.5 \text{mg/dm}^3$	Gaus.	0,117597	0,1654	61,3175	0,277	$1,033 \times 10^{-3}$
$P \leq 18 \text{mg/dm}^3$	Gaus.	0,156410	0,252815	196,8929	0,381	$2,573 \times 10^{-3}$
$\text{pH} \geq 5,2$	Gaus.	0.1036	0.1554	150,66	0,568	$2.155 \times 10^{-3}$
$\text{pH} \leq 5,4$	Gaus.	0,1639	0,1629	273,2903	0.203	$1,565 \times 10^{-3}$

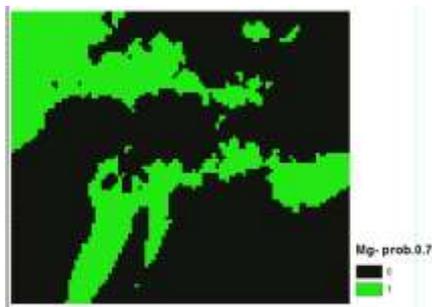
Pode-se notar na Tabela 1 que o IDE (Índice de Dependência Espacial) do Magnésio, Fósforo, pH e Cálcio foi baixo, demonstrando que há pouca dependência espacial para estas variáveis, de acordo com o espaçamento amostral adotado, porém, se percebe que estas variáveis podem apresentar-se clusterizadas com diferentes médias na área e por isso o estudo de mapas separadamente para cada nível mínimo e máximo de corte permite melhor conhecimento da distribuição espacial dos níveis das variáveis em estudo.

A partir dos valores dos parâmetros de semivariogramas foi gerados os mapas de Krigagem Indicativa com pontos de corte de acordo com a coluna (1) da Tabela 1. Para todas as variáveis foi feito dois semivariogramas para contemplar os dois pontos de corte de máximo e mínimo para o nível de suficiência e a partir disso foi gerado dois mapas de Krigagem Indicativa para cada variável, que em seguida, foi aplicado álgebra de mapas para se obter as regiões de intersecção desses dois mapas.

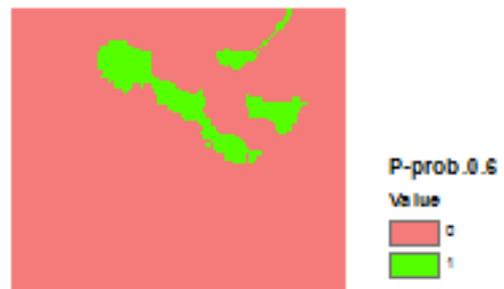
As Figuras (1,2, 3, 4, e 5), abaixo mostram os resultados obtidos com as álgebras de mapas para cada uma das variáveis estudadas.



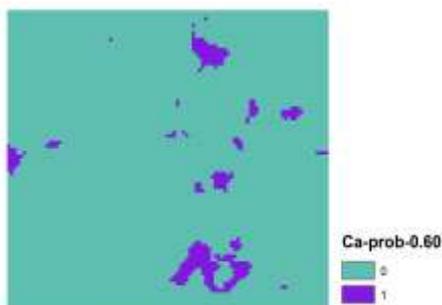
(a) Área com probabilidade maior ou igual a 0.6 de encontrar **pH** na faixa de 5,2 a 5,4.



(b) Área com probabilidade maior ou igual a 0.7 de encontrar **Mg** na faixa de 0.9 a 1.0 cmol<sub>c</sub>/dm<sup>3</sup>.



(c) Área com probabilidade mínima de 0.60 de encontrar **P** na faixa de 15.5 a 18 mg/dm<sup>3</sup>.



(d) Área com probabilidade maior ou igual a 0.6 de encontrar **Ca** na faixa de 1.9 a 2.25 cmol<sub>c</sub>/dm<sup>3</sup>.



(e) Área com probabilidade maior ou igual a 0.75 de encontrar **K** na faixa de 6 a 73 mg/dm<sup>3</sup>.

De acordo com os resultados obtidos nas análises, percebeu-se que os níveis de suficiências de Potássio estão subdividindo a área em dois grupos, de um lado apresenta-se probabilidade elevada de estar acima do nível mínimo e do outro abaixo do limite máximo e portanto a área que encontra-se na intersecção dos dois extremos, com probabilidade igual ou superior a 0.75 é muito pequena indicando

que a região tem descontrolo de Potássio. O Magnésio e o Cálcio apresentam-se em uma grande parte da área acima do limite mínimo mas em quase toda a área apresenta-se fora do intervalo entre o mínimo e máximo, com probabilidade inferior a 60%, ou seja, os níveis de suficiência para estas variáveis apresentam-se insatisfatórios e tanto o Cálcio e quanto o Magnésio apresentam-se clusterizados em locais distintos na área. O Fósforo apresenta-se em grande parte da área acima do limite mínimo do intervalo, porém, com probabilidade mínima de 60% a área que apresenta o teor de P dentro do intervalo de máximo e mínimo é muito pequena. E finalizando, temos com probabilidade igual ou superior a 0,6, o teor do pH apresenta-se fora dos limites de tolerância do cafeeiro e pode comprometer a produtividade, uma vez que o solo apresentou pH abaixo do limite mínimo indicado para o café e por isso sugere fazer a correção do solo de acordo com as probabilidades indicadas em cada sub-área.

#### **4. Conclusão**

De acordo com as informações dos mapas de Krigagem Indicativa associados a aplicação de álgebra de mapas é possível planejar a correção do solo de forma localizada, fazendo a reposição dos nutrientes nos locais em que há carência/excesso. A associação de mapas de Krigagem Indicativa com Álgebras de mapas permite sobrepor e comparar mapas e ter uma visão correlacionada dos locais onde o solo apresenta-se em condições de fertilidade.

## **Referências**

- [1] ALVES, Leonardo Corrêa, M.Sc., Universidade Federal de Viçosa, setembro de 2012. **Faixas de suficiência e níveis ótimos de atributos da fertilidade do solo para a cultura do café em Minas Gerais**. Orientadora: Herminia Emilia Prieto Martinez. Coorientador: Júlio César Lima Neves.
- [2] LANDIM, P.M.B. Análise estatística de dados geológicos. 2. ed. rev. e ampl. São Paulo: UNESP, 2003.
- [3] MACHADO, L. O., LANA, A. M. Q., LANA, R. M. Q., GUIMARÃES, E. C., FERREIRA, C. V. Variabilidade espacial de atributos químicos do Solo em áreas sob sistema de plantio convencional. **R. Bras. Brasileira de Solo**, 31:591-599, 2007.
- [4] MALAVOLTA, E. **Manual de Nutrição Mineral de Plantas**. São Paulo: Agronômica Ceres, 2006. 500 p.
- [5] MATIELLO, J. B.; SANTINATO, R.; GARCIA, A. W. R.; ALMEIDA, S. R.; FERNANDES, D. R. **Cultura de café no Brasil: Novo Manual de Recomendações**. Rio de Janeiro-RJ e Varginha-MG: MAPA/PROCAFÉ Fundação PROCAFÉ, 2005. 438 p.

# Distribuição odd log-logística-normal com uma aplicação em estatística experimental

Altemir da Silva Braga<sup>1 4</sup>

Edwin M. Marcos Ortega<sup>2</sup>

Gislaine Vieira Duarte<sup>3 4</sup>

**Resumo:** A distribuição normal é uma das mais destacada dentro da área da estatística, porém tal distribuição não se ajusta bem a dados que apresentam assimetria. Isto ocorre, porque a essa distribuição possui apenas os dois primeiros momentos. Nilton et al. (2014) estudaram a família de distribuições odd log-logística-g que possui três parâmetros: um de locação, um de escala e um de forma. Neste sentido, utilizou-se a distribuição normal como função base na distribuição odd log-logística-g e a ajustou para um conjunto de dados proveniente de um delineamento inteiramente casualizado. Foi utilizado o método da máxima verossimilhança no ajuste e os critérios *AIC*, *CAIC* e *BIC* para comparar os modelos. Após a modelagem estatística, a distribuição odd log-logística-normal ajustou-se melhor aos dados.

**Palavras-chave:** Ajuste; Assimetria; Comparação; Gama; Chances

## 1 Introdução

A estatística experimental tem por objetivo estudar dados provenientes de experimentos, contribuindo na pesquisa científica desde o planejamento até a interpretação dos resultados, conforme Banzatto e Kronka (2006). Dessa forma, a análise estatística passa a exercer um papel muito importante no contexto científico, devido suas técnicas serem utilizadas em quase todas as etapas da pesquisa.

Neste contexto, muitos estudos são realizados na área da estatística experimental com a finalidade de melhorar as análises e ajudar na interpretação dos dados. Dentre as distribuições de probabilidades utilizadas na área da experimentação a distribuição normal é uma das mais

---

<sup>1</sup>USP/ESALQ: Doutorando da Escola Superior de Agricultura Luiz de Queiros, Piracicaba-SP. Email: *altemir@usp.br*

<sup>2</sup>USP/ESALQ: Professor Titular da Escola Superior de Agricultura Luiz de Queiros, Piracicaba-SP. Email: *edwin@usp.br*

<sup>3</sup>USP/ESALQ: Doutoranda da Escola Superior de Agricultura Luiz de Queiros, Piracicaba-SP. Email: *gislainevduarte@gmail.com*

<sup>4</sup>FAPAC: Agradecimento à Fundação de Amparo à Pesquisa do Estado do Acre pelo apoio financeiro.

destacadas. Porém, essa distribuição não se ajusta bem aos dados que apresentam assimetrias, devido possuir apenas os dois primeiros momentos, ou seja, a média (parâmetro de locação) e o desvio-padrão (parâmetro de escala).

Por isso, muitos estudos são realizados com o objetivo de criar novas famílias de distribuições que acomodem, tanto a medida de assimetria, quanto a curtose dos dados. Conseqüentemente, tais famílias passam a ter mais parâmetros e exigir mais recursos computacionais. Pode-se encontrar na literatura algumas extensões da distribuição normal, como por exemplo, skew-normal, beta-normal, kumaraswamy-normal, a gama-normal e a odd log-logística-normal.

A família de distribuições odd log-logística-g foi proposta por Nilton et al. (2014) com o objetivo de criar uma nova família de distribuições que pudesse estender qualquer distribuição contínua. Os autores, neste estudo, demonstraram as propriedades estruturais, o comportamento da densidade para algumas distribuições base, como por exemplo, Weibull, Frechet, Half-normal, Log-logística e a log-normal, além de realizarem ajustes para dados reais com censuras. Dessa forma, dada uma função base  $G(y, \xi)$  com um vetor de parâmetros  $\xi$  pode-se obter uma extensão da odd log-logística-g, a uma nova distribuição com um parâmetro adicional de forma  $\alpha > 0$ , cuja distribuição foi definida por:

$$F(x|\mu, \sigma, \alpha) = \int_0^{\frac{G(y;\xi)}{\bar{G}(y;\xi)}} \frac{\alpha y^{\alpha-1}}{(1+y^\alpha)^2} dx = \frac{G(y;\xi)^\alpha}{G(y;\xi)^\alpha + \bar{G}(y;\xi)^\alpha}, \quad (1)$$

em que,

$$\alpha = \frac{\log \left[ \frac{G(y)}{\bar{G}(y)} \right]}{\log \left[ \frac{G(y)}{G(y)} \right]} \quad \text{and} \quad \bar{G}(y; \xi) = 1 - G(y; \xi).$$

Derivando a função distribuição  $F(x|\mu, \sigma, \alpha)$  tem-se a função de densidade da *odd log-logística-G* para a nova família dada por:

$$f(x|\mu, \sigma, \alpha) = \frac{\alpha g(y; \xi) \{G(y; \xi)[1 - G(y; \xi)]\}^{\alpha-1}}{\{G(y; \xi)^\alpha + [1 - G(y; \xi)]^\alpha\}^2}. \quad (2)$$

Dessa forma, conduziu-se este estudo, com o objetivo de ajustar e comparar a distribuição odd log-logística-normal com a distribuição normal a um conjunto de dados proveniente de um delineamento inteiramente casualizado. Após análise estatística a distribuição odd log-logística-normal ajustou-se melhor aos conjuntos de dados.

## 2 Metodologia

Neste estudo foi utilizado um conjunto de dados proveniente de um experimento de um delineamento inteiramente casualizado para avaliar o efeito da dose de um vermífugo (ml) no controle de um parasito (efeitos fixos). No delineamento constavam 5 tratamentos  $T_1$  e  $T_2$  - controles e  $T_3$ ,  $T_4$  e  $T_5$  foram uma nova droga com doses 5%, 10% e 15%) e 6 repetições. Os dados podem ser acessados na página do professor Euclides Malheiros Braga no endereço: <http://jaguar.fcav.unesp.br/euclides/>. Escolhendo o ano 2013 e a opção Estatística Experimental - PG em Ciência Animal (*UFERSA*) tem-se acesso ao arquivo *A\_DIC\_ex2.txt*.

No delineamento inteiramente casualizado o modelo estatístico tem a forma  $Y_{ij} = m + \tau_i + \xi_{ij}$ , em que  $i = 1, 2, \dots, I$ , sendo que  $I$  representa a quantidade de tratamento. Assim, para esse modelo a função de densidade da distribuição odd log-logística-normal é dada por:

$$f(y; \alpha, \sigma, m, \tau) = \frac{\alpha \left(\frac{1}{\sigma}\right) \phi \left[\frac{y-m-\tau_i}{\sigma}\right] \left\{ \Phi \left[\frac{y-m-\tau_i}{\sigma}\right] \left[1 - \Phi \left[\frac{y-m-\tau_i}{\sigma}\right]\right] \right\}^{\alpha-1}}{\left\{ \Phi \left[\frac{y-m-\tau_i}{\sigma}\right]^\alpha + \left[1 - \Phi \left[\frac{y-m-\tau_i}{\sigma}\right]\right]^\alpha \right\}^2},$$

fazendo  $\alpha = 1$ , a densidade tem a forma dada por,

$$f(y; \sigma, m, \tau) = \frac{1}{\sigma} \phi \left[ \frac{y - m - \tau_i}{\sigma} \right],$$

Para estimar os parâmetros da distribuição odd log-Logística-normal foi utilizado o método da máxima verossimilhança. sendo que para a obtenção das estimativas dos parâmetros  $(\alpha, \mu, \sigma)$  trabalhou-se com a função log verossimilhança (3) dada pela expressão:

$$\begin{aligned} l(\theta) &= n \log(\alpha) + \sum_{i=1}^n \left\{ \log \left[ \left( \frac{1}{\sigma} \right) \phi \left( \frac{x_i - \mu}{\sigma} \right) \right] \right\} + (\alpha - 1) \sum_{i=1}^n \left\{ \log \left[ \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right] \right\} \\ &\quad - 2 \sum_{i=1}^n \log \left\{ \left\{ \Phi \left( \frac{x_i - \mu}{\sigma} \right)^\alpha + \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^\alpha \right\} \right\} \\ &\quad + (\alpha - 1) \sum_{i=1}^n \left\{ \log \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right] \right\}. \end{aligned} \quad (3)$$

Na sequência foram obtidos os componentes do vetor escore,  $U(\theta)$  na seguinte ordem, primeiro  $U_\alpha(\theta)$ , depois  $U_\mu(\theta)$  e por último  $U_\sigma(\theta)$ . Assim, após derivar em relação  $\alpha$ ,  $\mu$  e  $\sigma$ , foi obtido que:

$$\begin{aligned}
U_\alpha(\theta) &= -2 \sum_{i=1}^n \left\{ \frac{\left\{ \Phi \left( \frac{x_i - \mu}{\sigma} \right)^\alpha \log \left[ \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right] + \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^\alpha \log \left\{ \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right] \right\} \right\}}{\left\{ \Phi \left( \frac{x_i - \mu}{\sigma} \right)^\alpha + \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^\alpha \right\}} \right\} \\
&\quad + \frac{n}{\alpha} + \sum_{i=1}^n \left\{ \log \left[ \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right] \right\} + \sum_{i=1}^n \left\{ \log \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right] \right\}. \\
U_\mu(\theta) &= \left( \frac{1}{\sigma} \right) \sum_{i=1}^n \left\{ \left( \frac{x_i - \mu}{\sigma} \right) \right\} + \left( \frac{\alpha - 1}{\sigma} \right) \sum_{i=1}^n \left\{ \frac{\phi \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 - 2\Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]}{\Phi \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]} \right\} \\
&\quad - 2 \sum_{i=1}^n \left\{ \frac{\left\{ \left( \frac{\alpha}{\sigma} \right) \phi \left( \frac{x_i - \mu}{\sigma} \right) \left\{ \Phi \left( \frac{x_i - \mu}{\sigma} \right)^{\alpha-1} - \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{\alpha-1} \right\} \right\}}{\left\{ \Phi \left( \frac{x_i - \mu}{\sigma} \right)^\alpha + \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^\alpha \right\}} \right\} \\
U_\sigma(\theta) &= \left( \frac{n}{\sigma} \right) + (\alpha - 1) \sum_{i=1}^n \left\{ \left\{ \left( \frac{1}{\sigma} \right) \phi \left( \frac{x_i - \mu}{\sigma} \right) \left( \frac{x_i - \mu}{\sigma} \right) \right\} \left\{ \frac{1 - 2\Phi \left( \frac{x_i - \mu}{\sigma} \right)}{\Phi \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]} \right\} \right\} \\
&\quad - 2 \sum_{i=1}^n \left\{ \frac{\left[ \alpha \left( \frac{1}{\sigma} \right) \phi \left( \frac{x_i - \mu}{\sigma} \right) \left( \frac{x_i - \mu}{\sigma} \right) \right] \left\{ \Phi \left( \frac{x_i - \mu}{\sigma} \right)^{\alpha-1} - \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{\alpha-1} \right\}}{\left\{ \Phi \left( \frac{x_i - \mu}{\sigma} \right)^\alpha + \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^\alpha \right\}} \right\}
\end{aligned}$$

### 3 Resultados e discussões

Na Tabela (2) estão as estimativas dos parâmetros para cada um dos modelos odd log-logística-normal (OLL-N) e normal (N). Dentre essas estimativas de máxima verossimilhança (EMVs) destacam -se os erros-padrão e os valores do Critério de Informação de Akaike (AIC), Critério da Informação Akaike(CAIC) e Bayesian Information Criterion (BIC). Quanto menor forem os valores destes critérios, melhor o ajuste. Pode-se, observar por meio dos erros-padrão que as estimativas do modelo odd log-logística-normal foram mais precisas e, além disso, pelos critérios (AIC) e (BIC) esse modelo explica melhor os dados. Este fato, foi validado e seus resultados estão apresentados na Tabela (1) que mostra o valor do teste de razão de verossimilhança e o nível descritivo. Deixando evidências de que existem diferenças estatísticas entre a distribuições odd log-logística-normal e a normal. Assim, pode-se afirmar que a distribuição odd log-logística-normal explica melhor os dados do que a distribuição normal.

Tabela 1: Teste da razão de verossimilhança.

Modelos	Hipóteses	Estatísticas $w$	$p$ -valor
OLL-N vs Normal	$H_0 : \alpha = 1$ vs $H_1 : H_0$ is false	3.950	0.0468

Tabela 2: Estimativas de verossimilhanças e as estatísticas AIC, CAIC e BIC

Modelos	Parâmetros	Estimativas	EP	AIC	CAIC	BIC
N	$m$	2477.132	99.088			
	$\tau_1$	-402.180	140.141			
	$\tau_2$	-1950.010	140.139	426.7	433.5	435.1
	$\tau_3$	-2320.570	140.139			
	$\tau_4$	-2385.700	140.134			
	$\sigma$	242.763	31.332			
OLL-N	$m$	2466.522	110.228			
	$\tau_1$	-405.010	146.886			
	$\tau_2$	-1949.320	139.539			
	$\tau_3$	-2310.007	131.505	424.7	433.7	434.5
	$\tau_4$	-2375.437	131.817			
	$\sigma$	1022.357	1343.898			
	$\alpha$	5.231	7.057			

## Referências

- [1] BANZATTO, A. D.; KRONKA, S. N. *Experimentação agrícola*. 4.ed. Jaboticabal:Funep, 2006.
- [2] NILTON, C. J.; CORDEIRO, G. M.; ORTEGA, E. M.M.. A new family of distributions: Theory and applications in survival analysis, 2014.

# Determinantes das arrecadações do seguro de responsabilidade civil no Brasil: uma análise econométrica no período de 2004 a 2011.

Pablo Cescon Portes<sup>1 3</sup>

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

**Resumo:** O seguro de responsabilidade civil tem como objetivo transferir as indenizações, relacionadas a responsabilidade civil, impostas judicialmente ao segurado para a seguradora. Com o crescimento do mercado desse tipo de seguro no Brasil, o presente trabalho tem por objetivo apresentar um modelo econométrico com variáveis independentes que possam ter relações com esse crescimento, são elas: número de processos judiciais por ano no Brasil, gasto das seguradoras com sinistro relacionados a responsabilidade civil e o PIB per capita. A relação entre as variáveis foi estudada por meio do uso de ferramentas estatísticas, mais precisamente, análise de regressões lineares simples e múltiplas. O modelo ajustado, para o PIB per capita, apresentou os melhores ajustes, um coeficiente de determinação ( $R^2$ ) de 0,8746, tal estatística denota que a variável explica a variável dependente em 87,46% dos casos, além disso, a estatística apresentou uma significância estatística inferior a 5%, confirmando a confiabilidade do ajuste. A regressão linear múltipla não apresentou boa qualidade nos ajustes e os sinais das variáveis independentes não ficaram de acordo com o esperado. Com isso, constatou-se que o PIB per capita foi a variável que mais se relacionou com o aumento do total de prêmio pago anualmente para as seguradoras com o seguro de responsabilidade civil. As outras variáveis não tiveram, estatisticamente e probabilisticamente, resultados significativos.

**Palavras-chave:** Seguro de responsabilidade civil, Influência seguro de responsabilidade civil.

## 1 Introdução

A noção de responsabilidade vem da origem da própria palavra, do latim *repondere*, responder a alguma coisa, ou seja, responder pelos seus atos danosos (STOCO, 2007). Nesse sentido responsabilidade civil é a obrigação de indenizar uma pessoa, a qual foi causada dano. Dessa maneira a vítima tem seus bens indenizados, usando a ideia de retornar ao que existia antes do dano (*Status quo ante*), e o causador do dano responde pelos seus atos (RODRIGUES, 2003). O seguro de responsabilidade civil visa transferir os riscos que uma pessoa tem de ter que indenizar outra pessoa, por conta de causar dano, por culpa ou dolo, para o segurador.

O mercado de seguro de responsabilidade civil aumentou muito do início dos anos 2000 até 2013. De acordo com a SUSEP, de 2004 a 2011 o total arrecadado com prêmio de seguro de responsabilidade civil duplicou. Sabendo disso o trabalho tenta achar variáveis que possam estar afetando e, conseqüentemente, aumentando esse mercado de seguros.

## 2 Objetivo

O objetivo geral deste trabalho é iniciar as investigações empíricas do processo de determinação das arrecadações do seguro de responsabilidade civil no Brasil. Por sua vez, o objetivo específico é identificar e analisar o efeito do número de processos judiciais por ano no Brasil, do gasto das seguradoras com sinistro relacionados a responsabilidade civil e do PIB per capita na arrecadação do seguro de responsabilidade civil no período de 2004 a 2011.

### **3 Metodologia**

#### **3.1 Variáveis analisadas**

Além da variável dependente, prêmio total pago em seguro de responsabilidade civil (dados retirados da SUSEP, 2014), foram selecionadas três variáveis independentes, as quais foram definidas a partir da análise da literatura existente sobre o assunto, são elas: aumento de processos judiciais, gastos das seguradoras com sinistros e PIB per capita. Os conceitos e descrição das variáveis utilizadas são apresentadas a seguir:

**a) Número de Processos Judiciais:** é sabido que o número de processos e ações judiciais no Brasil tem aumentado. A população brasileira está mais engajada no que diz respeito a processos em geral. De acordo com o CNJ (2012) (Conselho Nacional de Justiça), entre 2007 e 2008, os processos judiciais cresceram mais que a própria população, foram abertas 2,4 milhões de ações contra 2 milhões de nascidos vivos no mesmo ano. Ainda de acordo com CNJ (2012) o país atingiu a marca dos 90 milhões de ações em 2011.

Com esse aumento de processos judiciais as pessoas ficam mais preocupadas com a possibilidade e risco de danos à terceiros e, conseqüentemente, ter que arcar com as indenizações e se prejudicarem economicamente por conta disso. Devido a essa maior preocupação da população em geral, foi escolhida essa variável para o estudo.

**b) Gasto com Sinistro:** conforme a SUSEP (2012), um sinistro, é a concretização de um risco coberto. Caso não esteja coberto pelo contrato de seguro, é denominado evento danoso não coberto ou evento não coberto. No Seguro de Responsabilidade Civil, caracteriza-se pela atribuição, ao Segurado, da responsabilidade pela ocorrência de um evento danoso, causando danos a terceiros, atendidas as disposições do contrato. (Circular SUSEP 437/12).

Assim, a quantidade de capital desembolsada pelas seguradoras com sinistros, há de ter muita relação com a quantidade de capital concebida com prêmios de seguros de responsabilidade civil.

**c) PIB per capita:** conforme Voltaire Marensi (2009), seguro de responsabilidade civil faz parte dos seguros contra danos, que garantem o segurado contra riscos que ameaçam seus bens, ou seja, o seguro protege as pessoas de prejuízos econômicos. É cabível o raciocínio de que, quanto maior o poder econômico e quanto mais alto o valor de seus bens, mas a pessoa preza por sua integridade financeira. Por conta desse pensamento, o crescimento do PIB per capita da população brasileira foi analisada, para averiguar uma possível relação com o mercado de seguro de responsabilidade civil.

### 3.2 Tratamento dos Dados

Foi utilizada a análise de regressão linear simples, para cada variável separadamente, depois foi feito uso da análise de regressão múltipla, para verificar a relação de todas as variáveis, ao mesmo tempo. Dessa maneira pode-se analisar o quanto cada variável independente (PIB per capita, número de processos judiciais e número de empresas) está relacionada à variável dependente, prêmio total pago em responsabilidade civil.

De acordo com Levine et al (2008), a análise de regressão possibilita que se desenvolvam modelos para prever os valores de uma variável numérica com base no valor de outras variáveis. Também permite que se identifique o tipo de relação matemática entre as variáveis (forte ou fraca). Na regressão linear simples utiliza-se apenas uma variável independente, seguindo a seguinte equação de primeiro grau, onde  $Y$  é a variável dependente,  $X$  é a variável independente,  $\beta_0$  é o intercepto da equação e  $\beta_1$  o coeficiente de  $X$ :  $Y = \beta_0 + X\beta_1$

A regressão linear múltipla faz uso de várias variáveis independentes que se relacionam em forma de uma equação de primeiro grau com uma única variável dependente, logo sua equação pode ser expressa dessa maneira:  $Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \dots + X_n\beta_n$

Foram usados valores de âmbito nacional, com uma série temporal de 2004 a 2011 anuais para todas as variáveis. O tratamento dos dados foram todos feitos no software estatístico R-Project.

### 4 Resultado e discussões

As estatísticas descritivas das variáveis podem ser vistas na Tabela 1.

Tabela 1: Estatísticas descritivas dos dados.

	média	mediana	desvio-padrão	variância	assimetria	curtose
Prêmios total	427,125	421,5	117,0477	12554,6964	0,588028	-0,94266
Processos	23,7	25,15	2,107809	4,337143	-0,388101	-1,56999
Sinistros	140,375	135	25,55631	653,125	0,962702	-0,37489
PIB (1%)	9713,375	9817,5	663,1918	439823,41	-0,62494	-1,11391
PIB (Geral)	785,4363	797,59	80,70607	6513,4702	-0,325519	-1,47733

Na Tabela 2 apresenta-se os dados das regressões lineares simples, construídas separadamente para cada variável independente.

Tabela 2: Regressões lineares simples.

	g.l.	intercepto	coeficiente	Valor-P	valor F	F de significação	R ajustado
Processos	6	-0,8267	2,5086	0,00279	23,745	0,002786	0,7674
Sinistros	6	0,4157	1,0284	0,059	5,4097	0,05897	0,3865
PIB (1%)	6	-10,2753	3,2343	0,00242	25,12	0,002423	0,7751
PIB (Geral)	6	-3,9438	2,2683	0,000405	49,822	0,0004048	0,8746

Tanto para Levine et al (2008), quanto para Draper e Smith (1998), os níveis de significância aceitos, para o teste F e o teste t, são de 5%, dessa maneira, adota-se o nível de significância de 5% para ambos os casos (F de significação e valor-P).

De acordo com a Tabela 1, as variáveis “PIB per capita da população 1% mais rica” e “PIB per capita geral do Brasil” apresentam os melhores números de uma regressão linear simples. Ambos têm o valor-P e o F de significação, bem menores que o nível de significância de 5%. Além disso, essas variáveis possuem o R ajustado maior que 0,70, ou seja, mais que 70% dos casos podem ser explicados por essas variáveis (77,51% e 87,46%). Logo podemos concluir que os modelos de regressão construídos, separadamente, para essas variáveis são os mais adequados.

A variável “processos judiciais” tem o F de significação e o valor-P menores que um nível de significância de 5%, além disso, tem seu R ajustado maior que 0,7, ou seja, o modelo ajustado é totalmente aceitável. A variável “total pago em sinistros” apresentou o pior modelo ajustado, com um R ajustado de 0,3865, muito menor que os Rs ajustados dos outros modelos. Além disso apresentou o F de significação e o valor-P maiores que um nível de significância de 5%, o que para maioria das pesquisas estatísticas é rejeitável.

Analisando a regressão linear múltipla, tem-se que o valor-P das variáveis é muito alto, muito maior que 0,05, os coeficientes de duas das variáveis independentes estão negativos, o que não faz sentido, levando em conta que ambos os valores aumentam junto com variável dependente. O R ajustado e o F de significação estão dentro do padrão, porém levando em conta a multicolinearidade achada no modelo, esta regressão é inadequada. Por isso o trabalho não analisará o modelo ajustado múltiplo.

## **5 Considerações Finais**

O trabalho mostrou a crescente procura por seguros de responsabilidade civil no Brasil, e as variáveis que podem estar afetando e se relacionando esse crescimento. A princípio acreditava-se que os processos judiciais teriam a maior relação com os seguros de responsabilidade civil, devido ao seu crescimento nos anos observados. Além disso, o trabalho expôs a ideia de que os processos, relacionados a responsabilidade civil, têm aumentado a necessidade de indenizar a vítima e em contrapartida diminuído a importância da culpa, o que remete a pensar numa relação forte entre o aumento de processos e o aumento de seguros de responsabilidade civil.

Porém, a análise estatística feita no trabalho, mostrou que essa relação não foi a mais forte. Apesar disso o modelo ajustado encontrado, se manteve dentro dos padrões de uma regressão linear simples, o que para o trabalho é um modelo aceitável.

Por outro lado, foi mostrado que o PIB per capita, tanto o PIB per capita geral do Brasil, como o PIB da população 1% mais rica do Brasil tem uma relação estatística mais forte que as outras variáveis. Dessa maneira pode-se concluir que os salários da população afetam o mercado de seguro

de responsabilidade civil. Isso ocorre pela necessidade que a população, agora mais rica, tem de resguardar seu dinheiro, várias pesquisas feitas nessa área mostram essa necessidade.

O gasto das seguradoras com sinistros, relacionados a responsabilidade civil, mostrou ser a única variável que não se ajustou bem a regressão linear simples. Estatisticamente os números mostraram não haver uma relação, considerada aproveitável. Conclui-se que o PIB per capita foi a variável mais relevante apresentada, pois foi a que melhor respondeu estatisticamente a regressão linear. Para um trabalho futuro pode-se fazer uma análise mais profunda em cima dos fatos apresentados no trabalho, por exemplo, como exatamente o PIB influencia diretamente esse mercado?

Por fim, é importante ressaltar algumas limitações deste trabalho, como a questão inerente a metodologia, ou seja, a regressão linear, pois esta trabalha com as variáveis em um caráter probabilístico. Outro ponto fraco do trabalho é a dificuldade para encontrar os dados brutos, sendo o número de processos judiciais por ano no Brasil o dificultoso para achar. Além disso, é escassa a literatura sobre o enfoque dado neste trabalho, há muita coisa escrita sobre responsabilidade civil e seus tipos de seguro, porém, trabalhos sobre como esse mercado é influenciável por variáveis não foram encontrados.

## Referências

- [1] CONSELHO NACIONAL DE JUSTIÇA. *Relatório Justiça em Números 2010*. Disponível em: <<http://www.cnj.jus.br/programas-de-a-a-z/eficiencia-modernizacao-e-transparencia/pj-justica-em-numeros/relatorios>>. Acesso em: 24 de abril de 2013.
- [2] LEVINE, D. M.; STEPHAN, D. F.; KREHBIEL, T. C. BERENSON, Mark L. *Estatística: Teoria e Aplicação*. LTC. 5ª ED.2008.
- [3] MARENSE, V. *O Seguro no Direito Brasileiro*. Lumen Juris. 9ª ED. 2009.
- [4] RODRIGUES, S. *Direito Civil: Responsabilidade Civil*. Saraiva. 20ª ED. 2003.
- [5] SANTOS, P. P. S. *Responsabilidade Civil: origem e pressupostos gerais*. Disponível em: <[http://www.ambito-juridico.com.br/site/?n\\_link=revista\\_artigos\\_leitura&artigo\\_id=11875](http://www.ambito-juridico.com.br/site/?n_link=revista_artigos_leitura&artigo_id=11875)>. Acesso em: 04 Julho 2014.
- [6] STOCO, R. *Tratado de Responsabilidade Civil: Doutrina e Jurisprudência*. Revista dos Tribunais. 8ª ED. 2011.
- [7] SUSEP. *Circular SUSEP Número 437 de 14 de junho de 2012*. Disponível em: <<http://www2.susep.gov.br/bibliotecaweb/docOriginal.aspx?tipo=1&codigo=29548>>. Acesso em: 05 Junho 2014.

# Determinantes macroeconômicos da taxa Selic: um estudo econométrico no período de 1994 a 2014.

Maria Eugênia de Oliveira Batista<sup>1 4</sup>

Gabriel Rodrigo Gomes Pessanha<sup>2 4</sup>

Jennifer Elis Pires<sup>3 4</sup>

**Resumo:** A taxa de juros Selic (Sistema Especial de Liquidação e Custódia) é considerada a taxa de juros básica no curto prazo da economia brasileira. Com o propósito de avaliar quais variáveis influenciam essa taxa, o trabalho utiliza métodos econométricos para analisar se as variáveis macroeconômicas (PIB, taxa de câmbio, inflação, índice Ibovespa) e as duas variáveis dummy (abertura de 1999 e crise de 2008) escolhidas, possuem influência sob a Selic no período entre julho de 1994 a fevereiro de 2014. A análise de regressão constatou que 84% das variações da taxa Selic são explicadas pelas variáveis estudadas. A abertura de 1999, inflação e o PIB apresentaram os resultados mais significativos, e o índice Ibovespa não obteve resultado expressivo.

**Palavras-chave:** *Selic, Regressão múltipla.*

## 1. Introdução

O Brasil possui como taxa básica de juros da economia, a taxa Selic (Sistema Especial de Liquidação e Custódia), responsável por regular diariamente os preços unitários dos títulos públicos, é a taxa utilizada pelo Banco Central a fim de vender e comprar esses títulos com o objetivo de estruturar sua política monetária.

Após 1994, início do Plano Real, as taxas de juros Selic aumentaram significativamente com o propósito de conter a taxa de inflação (HARFUCH, 2008). A partir de então, a taxa de juros também se tornou o instrumento mais significativo do Banco Central para a determinação das metas de inflação. Pode-se dizer que o grande problema enfrentado pelo governo afim de conseguir um equilíbrio macroeconômico e um desenvolvimento econômico está relacionada com as altas taxas de juros de curto prazo que o Brasil possui (BRESSER-PEREIRA, 2007).

## 2. Objetivo

---

<sup>1</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha - MG. Email: [mag\\_batista@hotmail.com](mailto:mag_batista@hotmail.com)

<sup>2</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha - MG. Email: [gabriel.pessanha@unifal-mg.edu.br](mailto:gabriel.pessanha@unifal-mg.edu.br)

<sup>3</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha - MG. Email: [jenni.elis@hotmail.com](mailto:jenni.elis@hotmail.com)

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

Com o propósito de analisar os valores da taxa Selic, o presente estudo teve como objetivo averiguar, através da regressão linear múltipla, se as variáveis macroeconômicas (PIB, inflação, taxa de câmbio, índice IBOVESPA e duas variáveis *dummy* – correspondentes à crise de 2008 e à abertura do câmbio, em 1999, período em que houve a troca de câmbio fixo para câmbio flutuante), possuem alguma influência no processo de formação da taxa SELIC, no período correspondente a julho de 1994, início do Plano Real, a fevereiro de 2014.

### 3. Metodologia

Os dados correspondentes à variável dependente Selic e às independentes PIB, inflação, taxa de câmbio e índice Ibovespa foram coletados através do site IPEADATA (<http://www.ipeadata.gov.br/>). Além disso, as variáveis *dummy* foram incorporadas como valor 0 até seu acontecimento, e a partir daí receberam valor 1 até o final do período. O período de pesquisa compreende-se entre julho de 1994, início do Plano Real, à fevereiro de 2014. Para a análise de regressão foi utilizado o *software Gretl*.

A análise de regressão linear múltipla foi utilizada como principal método de estudo. Esta análise diz respeito ao estudo de uma variável dependente, em relação a uma ou mais variáveis independentes, visando estimar e/ou prever o valor médio populacional da variável dependente em termos dos valores conhecidos ou fixados (em amostragens repetidas) das independentes. (GUJARATI e PORTER, 2011).

A equação geral referente à regressão linear múltipla pode ser expressa como:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + u \quad (1)$$

Onde,

$y$  = observações da variável dependente ou variável resposta;

$x$  = variáveis independentes existentes;

$\beta_0$  = intercepto;

$\beta_j$  = mudança esperada em  $y$  devido ao aumento de uma unidade em  $x_j$ , estando as outras variáveis fixas, ou seja, são os parâmetros da população;

$u$  = erro padrão

Primeiramente é realizado um teste de hipótese, considerado uma regra usada para decidir se uma hipótese estatística deve ser rejeitada ou não. O objetivo deste teste é decidir se uma hipótese sobre determinada característica da população é ou não apoiada pela evidência obtida de dados amostrais. E eles são os primeiros estudos realizados para a verificação da validade do modelo.

A hipótese formulada e testada no decorrer deste trabalho refere-se à influência ou não das variáveis macroeconômicas na taxa Selic. Especificamente com relação as variáveis independentes, espera-se um resultado positivo com as variáveis inflação, câmbio e PIB. E em relação às outras

variáveis, é esperado que elas possuam relação negativa com a variável dependente, assim, com o aumento nas variáveis independentes, espera-se uma diminuição na taxa Selic.

#### 4. Análise dos Resultados

Na pesquisa, algumas variáveis como a taxa Selic, taxa de câmbio, PIB e Ibovespa passaram por um processo de logaritmização como forma de tornar os dados lineares, e, com base nestes dados foi possível a elaboração dos resultados.

Na Tabela 1 apresenta-se os coeficientes de correlação das variáveis estudadas. Tais estatísticas foram importantes para a verificação da presença de multicolinearidade entre as variáveis estudadas.

Tabela 1: Resultado da análise de correlação

	SELIC	CÂMBIO	PIB	IBOVESPA	INFLAÇÃO	Crise 2008	Abertura 1999
SELIC	1						
CÂMBIO	-0,47654529	1					
PIB	-0,8956132	0,508943213	1				
IBOVESPA	0,0071811	0,047764876	0,00674328	1			
INFLAÇÃO	0,471267538	-0,20191005	-0,37910345	0,1175701	1		
Crise 2008	-0,68550621	0,100563533	0,76380164	-0,04549447	-0,1759465	1	
Abertura 1999	-0,68801152	0,852947055	0,70882729	0,07796409	-0,2700632	0,33368329	1

Fonte: Elaborado pela autora (2014).

Os resultados negativos (<0) existentes na tabela nos revelam uma relação negativa entre as variáveis, ou seja, movimento de alta na variável dependente são acompanhadas por movimento de baixa da variável independente. Conseqüentemente, quando os resultados são positivos (>0), existe uma relação positiva entre as variáveis, ou seja, se a variável dependente aumenta, a independente também irá aumentar.

Apesar de existirem alguns valores elevados na análise de correlação, foi realizado um teste de colinearidade, mostrando que esses valores não influenciaram na análise dos resultados. O valor mínimo possível da colinearidade é 1,0, e os valores > 10,0 podem indicar um problema de colinearidade. Os valores de colinearidade das variáveis são: a taxa de câmbio obteve 4,320 como valor, o PIB 6,131, Ibovespa 1,039, a variável crise de 2008 3,489, e a abertura 1999 5,954. Portanto, pode-se concluir que não ocorreram problemas de colinearidade, e não houve nenhuma influência na explicação dos dados.

Os resultados encontrados a partir do ajuste do modelo de regressão linear múltipla, que estimou o comportamento da taxa Selic em relação às variáveis independentes apresentaram-se estatisticamente significativos. O coeficiente de determinação do modelo ( $R^2$ ) foi de,

aproximadamente, 84%, com uma significância inferior a 1% (teste F), ou seja, as variações da SELIC são explicadas pelas variações das variáveis independentes em 84% dos casos, conforme a Tabela 2.

Tabela 2: Resultados da análise de regressão

<i>Variáveis</i>	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Teste-t</i>
Interseção	2,641699612	0,246656269	10,710044***
CÂMBIO	0,129551045	0,072358395	1,790407*
PIB	-0,479560138	0,050345807	-9,525324***
IBOVESPA	0,000846627	0,000618998	1,367738
INFLAÇÃO	0,061585387	0,009199564	6,694381***
Crise 2008	-0,047430246	0,02298641	-2,063403**
Abertura 1999	-0,124061311	0,032186774	-3,854418***
R- Quadrado			0,8447475***

Fonte: Elaborado pela autora (2014).

Legenda: \* significativo à 10% \*\* significativo à 5% \*\*\* significativo à 1%.

Partindo para análise individual das variáveis percebe-se que a única variável não expressiva dentro do modelo foi o Ibovespa, pois não obteve significância estatística. Com relação ao PIB, pode-se afirmar existe uma relação inversa com a variável dependente, em outras palavras, um aumento no PIB de 10% acarretaria uma diminuição de, aproximadamente, 4,8% na taxa SELIC, tal estatística é significativa ao nível de 1%. Comportamento semelhante a este se observa nas variáveis *dummies* estudadas (Crise de 2008 e abertura de 1999), ambas variáveis apresentaram uma relação inversa com a taxa SELIC. Já as variáveis Câmbio e Inflação apresentaram uma relação direta com a variável dependente estudada, ou seja, um aumento de 10% nas variáveis independentes acarretaria uma elevação na taxa Selic na ordem de 1,2% e 0,6%, respectivamente, tais estatísticas são confirmadas pelo teste t com um nível de 10% para a variável Câmbio e 1% para a variável Inflação.

## 5. Considerações finais

Com o objetivo de estudar as influências de algumas variáveis macroeconômicas sobre o processo de formação da taxa Selic utilizou-se a metodologia de regressão linear múltipla.

Inicialmente, examinaram-se as séries estudadas para verificar se as mesmas atendiam os pressupostos teóricos da regressão. Para tornar mais lineares os valores de algumas das variáveis como taxa Selic, PIB, taxa de câmbio e índice Ibovespa, estes passaram por um processo de logaritmização, para que assim, fosse possível a realização do método.

Os resultados obtidos apresentaram valores expressivos. O coeficiente de determinação do modelo ( $R^2$ ) foi de, aproximadamente, 84%, com significância estatística inferior a 1%, ou seja, as variações da SELIC são explicadas pelas variações das variáveis independentes em 84% dos casos.

Partindo para a análise individual das variáveis dependentes sob a independente, a única variável que não obteve resultado significativo foi o índice Ibovespa, pelo fato de obter um “valor-P” alto, e rejeitando a hipótese  $H_0$  de significância. Os demais valores das variáveis possuíram resultados com níveis de significância variando entre 10%, 5% e 1%. A taxa de câmbio apresentou uma significância de 10%, a variável dummy Crise de 2008 apresentou um nível de significância de 5%. E as demais variáveis apresentaram resultados significativos à 1%.

Por fim, pode-se concluir que as variáveis incluídas no modelo explicaram grande parte das variações da taxa SELIC no período estudado, todavia sugere-se para estudos futuros a inclusão de outras variáveis a fim de se aumentar o poder de explicação.

## Referências

- [1] BRESSER-PEREIRA, L.C. *Macroeconomia da estagnação: crítica da ortodoxia convencional no Brasil pós-1994*. São Paulo: editora 34, 2007.
- [2] GUJARATI, D.N.; PORTER, D.C. *Econometria Básica*, 5 ed. Porto Alegre: AMGH, 2011.
- [3] HARFUCH, Leila. *Determinantes da taxa de juros nominal e sua relação com a taxa de câmbio no Brasil no período de 1990 a 2006*. Piracicaba, 2008.

# Uma Análise de Desempenho de Gráficos de Controle Multivariados

Rodrigo Luiz Pereira Lara<sup>1 4</sup>

José Ivo Ribeiro Júnior<sup>2 4</sup>

Rafael Lédo Rocha de Oliveira<sup>3 4</sup>

**Resumo:** A melhoria e o monitoramento da qualidade são preocupações e necessidades para muitas organizações. Dentre as ferramentas que compõem o controle estatístico do processo, os gráficos de controle são simples e eficientes no monitoramento da média e da variabilidade de diversas características avaliadas. No presente trabalho são apresentados os gráficos de controle multivariados  $T^2$  de Hotelling, MEWMA, MCUSUM e ELR, com  $ARL IC = 200$ . Para cada um deles, simularam-se 20.000 processos bivariados sob controle baseados em 500 subgrupos racionais com observações individuais. A partir dos limites superiores de controle selecionados para compor os gráficos, modificou-se o vetor de médias do processo de modo a verificar em qual subgrupo racional, em média, o gráfico detecta o descontrole. O gráfico de controle MEWMA apresentou melhor desempenho, enquanto que o ELR e MCUSUM também apresentaram atuações satisfatórias.

**Palavras-chave:** *Alarme falso, Planejamento, ARL IC.*

## 1 Introdução

Dada a volumosa pesquisa em várias áreas de gráficos de controle univariados, a pesquisa em gráficos de controle multivariados talvez esteja em atraso (ZHANG, LI & WANG, 2010). Isso é provável, pois em várias aplicações industriais, a qualidade dos produtos é frequentemente relacionada a várias características de qualidade correlacionadas. Vários autores tais como Moraes et al. (2014) também apontaram que os gráficos de controle multivariados são uma importante área de pesquisa para o novo século.

O primeiro gráfico de controle multivariado foi proposto por Hotelling (1947), caracterizado por ser eficaz em detectar grandes mudanças em um processo, além de possuir um critério de decisão relativamente simples que se baseia apenas na mais recente amostra produzida. Posteriormente foram propostos os gráficos da Soma Acumulada Multivariada (*Multivariate Cumulative Sum* – MCUSUM) de Crosier (1988) e da Média Móvel Exponencialmente Ponderada Multivariada (*Multivariate*

---

<sup>1</sup> DET – UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [rodrigo.lara@ufv.br](mailto:rodrigo.lara@ufv.br)

<sup>2</sup> DET – UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [jivo@ufv.br](mailto:jivo@ufv.br)

<sup>3</sup> IFNMG: Instituto Federal do Norte de Minas Gerais, Januária-MG. Email: [rafaelrocha.est@gmail.com](mailto:rafaelrocha.est@gmail.com)

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

*Exponentially Weighted Moving Average* – MEWMA) de Lowry et al. (1992), cujos critérios de decisão se baseiam no histórico do processo com sensibilidade em detectar pequenas mudanças no vetor de médias. Recentemente, Zhan, Li e Wang (2010) propuseram o gráfico de controle ELR que monitora simultaneamente o vetor de médias e a matriz de covariâncias de um processo multivariado. No presente trabalho é realizado uma comparação de desempenho dos gráficos de controle multivariados  $T^2$  de Hotelling, MEWMA, MCUSUM e ELR, por meio de um monitoramento do vetor de médias de processos simulados.

Quando há  $p$  características a serem analisadas, o monitoramento é feito sobre um vetor de médias. Daí deseja-se testar se o vetor de médias  $\mu$  é estatisticamente igual ao vetor de médias  $\mu_0$  das características de qualidade, a fim de verificar se o processo está sob controle estatístico para todas elas.

As estatísticas dos gráficos de controle  $T^2$  de Hotelling, MEWMA, MCUSUM e ELR podem ser visualizadas em Hotelling (1947), Lowry *et al.* (1992), Crosier (1988) e Zhang, Li e Wang (2010), respectivamente. Foram definidos  $k = 1$  e  $r = 0,2$  para os fatores presentes nos gráficos de controle MCUSUM e MEWMA respectivamente.

Além desta introdução, o trabalho possui mais quatro seções principais. No item seguinte são apresentados os objetivos propostos. Em seguida tem-se a metodologia empregada; na quarta Seção, por meio de simulações de processos, são apresentados os resultados e discussões. Por fim, tem-se a conclusão e algumas considerações.

## 2 Objetivo

Dentre os gráficos de controle multivariados de Hotelling, MEWMA, MCUSUM e ELR, o objetivo do presente trabalho é determinar quais gráficos de controle são eficazes na detecção de mudança no vetor de médias de um processo. Mais especificamente, pretende-se verificar quantos subgrupos racionais são gastos, em média, por cada um dos gráficos de controle de modo a detectar a mudança do processo.

## 3 Metodologia

### 3.1 Processo Sob Controle Estatístico

Para cada um dos quatro gráficos de controle multivariados, simularam-se 20.000 processos bivariados sob controle com 500 subgrupos racionais cada e para diferentes valores de  $h$  (limite superior de controle), de modo a verificar em qual valor de  $h$  ocorre um comprimento média da sequência (*ARL IC*) igual a 200. Em outras palavras, deseja-se determinar para qual valor de  $h$  ocorre

alarme falso, em média, no 200º subgrupo racional ( $ARL IC = 200$ ). O vetor de médias e a matriz de covariâncias do processo considerado sob controle estão especificados na Equação 3.3.

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (3.1)$$

### 3.2 Processo Fora de Controle Estatístico

A partir dos valores de  $h$  que proporcionaram  $ARL IC = 200$ , simulou-se o descontrole no processo, conforme descrito na seção a seguir. Em função dos limites superiores de controle selecionados, modificou-se o vetor de médias do processo de modo a verificar em qual subgrupo racional, em média, cada gráfico detecta o descontrole. Realizaram-se 60.000 simulações de processos que foram monitorados pelos respectivos gráficos. Cada processo já foi inicializado fora de controle somente para o vetor de médias, como segue:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0,5 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (3.2)$$

Todas as simulações de processos sob e fora de controle foram realizadas utilizando-se o software R.

Tabela 4.1 - Limites superiores de controle obtidos por simulação em função do comprimento médio da sequência para os gráficos de controle  $T^2$  de Hotelling, MCUSUM, MEWMA e ELR

$T^2$ de Hotelling		MCUSUM ( $k = 1$ )		MEWMA ( $r = 0,2$ )		ELR	
$h$	$ARL IC$	$h$	$ARL IC$	$h$	$ARL IC$	$h$	$ARL IC$
11,2	176,45	3,15	175,59	10,9	190,59	1,860	196,46
11,3	179,33	3,20	182,11	11,0	194,17	1,861	197,51
11,4	183,11	3,25	188,71	11,1	196,70	1,862	197,55
11,5	186,46	3,30	191,48	<b>11,2</b>	<b>199,74</b>	1,863	197,71
11,6	189,71	3,35	196,96	11,3	201,03	<b>1,864</b>	<b>200,35</b>
11,7	192,61	<b>3,40</b>	<b>202,45</b>	11,4	202,43	1,865	201,61
11,8	196,98	3,45	208,07	11,5	203,72	1,866	201,25
11,9	196,92	3,50	210,11	11,6	207,33	1,867	198,64
<b>12,0</b>	<b>200,07</b>	3,55	214,08	11,7	209,08	1,868	201,92
12,1	201,65	3,60	219,56	11,8	210,83	1,869	198,64

$h$  = limite superior de controle,  $k$  = valor de referência,  $r$  = parâmetro de suavização.

## 4 Resultados e Discussões

### 4.1 Processo Sob Controle Estatístico

Na Tabela 4.1 é apresentado alguns valores obtidos de  $h$  em função do comprimento médio da sequência para um processo sob controle. Observe que quanto maior o valor de  $h$ , mais subgrupos racionais são necessários, em média, para a sinalização de um alarme falso, ou seja, a sinalização de um descontrolo quando na verdade o processo está operando sob sua devida média. As linhas em negrito da Tabela 1 são aquelas onde obteve-se  $ARL_{IC}$  próximo a 200 e, conseqüentemente, o valor de  $h$  associado, para os respectivos gráficos de controle.

Portanto, a partir de tais escolhas de  $h$ , todos os gráficos de controle estão em condições iguais para serem comparados realizando-se uma mudança no vetor de médias do processo.

### 4.2 Processo Fora de Controle Estatístico

Em média, o gráfico de controle MEWMA necessitou de 8,94 subgrupos racionais para detectar o descontrolo, sendo considerado o mais eficaz dentre os quatro gráficos utilizados nesse trabalho (Figura 4.1).

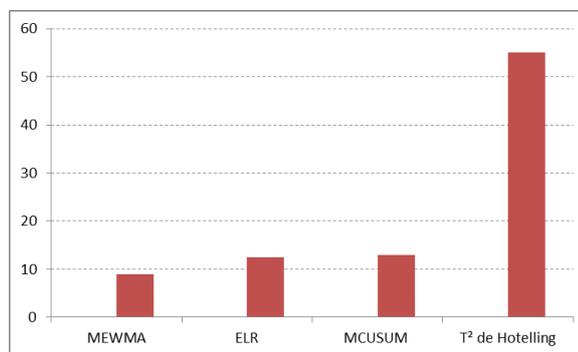


Figura 4.1 - Quantidade média de subgrupos racionais necessários para a detecção de mudança da média em processos bivariados

Os gráficos de controle ELR e MCUSUM apresentaram desempenhos satisfatórios. Ambos levaram, em média, uma quantidade de subgrupos racionais entre 12 e 13 para detectarem o descontrolo. O gráfico  $T^2$  de Hotelling foi o que apresentou atuação pior, uma vez que seu critério de decisão leva em consideração cada amostra separadamente, ao contrário dos demais gráficos que armazenam um histórico do processo.

Na Figura 4.2 é apresentado uma das 60.000 simulações de processos fora de controle estatístico monitorado pelo gráfico MEWMA. Nota-se que para esse exemplo, o gráfico detectou a mudança de média no oitavo subgrupo racional.

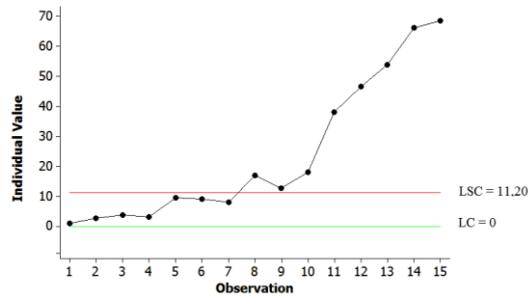


Figura 4.2 - Processo bivariado fora de controle estatístico monitorado pelo gráfico de controle MEWMA.

## 5 Conclusões

O gráfico de controle MEWMA foi o que apresentou o melhor desempenho. É importante observar que a escolha de  $r = 0,2$  contribuiu para a rápida detecção de pequenas mudanças no vetor de médias do processo.

Os gráficos de controle ELR e MCUSUM ( $k = 1$ ) apresentaram desempenhos bons. Embora o ELR não tenha sido o melhor, deve-se lembrar que ele também é capaz de monitorar variabilidade do processo, situação que não foi abordada nesse trabalho.

O gráfico de Hotelling foi o que apresentou o pior desempenho, uma vez que seu critério de decisão leva em consideração cada amostra separadamente, ao contrário dos demais gráficos que armazenam um histórico do processo.

## Referências

- [1] CROSIER, R. B. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*. vol. 30, n. 3, p. 291-303, 1988.
- [2] HOTELLING, H. Multivariate quality control. *Techniques of statistical analysis*. New York: Mc Graw Hill, p. 111-184, 1947.
- [3] LOWRY, C. A.; WOODALL, W. H.; CHAMP, C. W.; RIGDON, S. E. A multivariate exponentially weighted moving average control chart. *Technometrics*. vol. 34, n. 1, 1992.
- [4] MORAES, D. A. O; OLIVEIRA, F. L. P.; QUININO, R. C.; DUCZMAL, L. H. Self-oriented control charts for efficient monitoring of mean vectors. *Computers & Industrial Engineering*, vol. 75, p. 102-105, 2014.
- [5] ZHANG, J.; LI, Z.; WANG, Z. A multivariate control chart for simultaneously monitoring process mean and variability. *Computational Statistics and Data Analysis*, vol. 54, 2010.

# Análise de Arquétipos na avaliação da movimentação de jogadores de futebol

José Márcio Martins Júnior<sup>1 6</sup>

Elcio do Nascimento Chagas<sup>2 6</sup>

Eric Batista Ferreira<sup>3 6</sup>

Denismar Alves Nogueira<sup>4 6</sup>

Daniel Furtado Ferreira<sup>5 6</sup>

**Resumo:** A Análise de Arquétipos é uma técnica multivariada utilizada para reduzir a dimensão de dados por meio de combinações lineares. Como a Análise de Arquétipos utiliza principalmente extremos da massa de dados, este trabalho apresenta uma forma de utilizar esta técnica para avaliar a movimentação de um jogador ou de um grupo de jogadores de futebol, considerando as coordenadas geográficas (x,y) como variáveis originais. Para isso, foram utilizados dados fictícios gerados a partir de um esquema tático semelhante ao utilizado pela seleção brasileira durante a Copa do Mundo da FIFA, ocorrido no Brasil, em 2014. Resultados mostram que esta forma de utilizar a Análise de Arquétipos pode ser útil para técnicos de futebol, ao analisar o desempenho de seus times, após uma partida. Além disso, tal análise pode ser feita tendo-se como foco um jogador ou um grupo de jogadores (em particular, excetuando-se apenas o goleiro).

**Palavras-Chave:** *Análise de Arquétipos, Análise multivariada, Estatística no Esporte.*

## 1 Introdução

A Análise de Arquétipos (AA) foi introduzida por Cutler e Breiman (1994). É uma técnica multivariada utilizada com o propósito de simplificar a estrutura de covariâncias, sendo utilizada para reduzir a dimensão de dados por meio de combinações lineares dos seus elementos mais representativos. Os arquétipos são selecionados pela minimização da soma de quadrados de resíduos (SQR) ao representar cada observação como uma combinação dos arquétipos ou como um dos arquétipos (também denominado arquétipo puro) e estão na fronteira do fecho convexo dos dados. Portanto, são geralmente valores extremos que melhor representam os dados.

---

<sup>1</sup>ICEX - UNIFAL-MG. E-mail: [jmmjunifal@gmail.com](mailto:jmmjunifal@gmail.com)

<sup>2</sup>IFES - Campus de Alegre. E-mail: [enchagas@ifes.edu.br](mailto:enchagas@ifes.edu.br)

<sup>3</sup>ICEX - UNIFAL-MG. E-mail: [eric.ferreira@unifal-mg.edu.br](mailto:eric.ferreira@unifal-mg.edu.br)

<sup>4</sup>ICEX - UNIFAL-MG. E-mail: [denismar.nogueira@unifal-mg.edu.br](mailto:denismar.nogueira@unifal-mg.edu.br)

<sup>5</sup>DEX - UFLA. E-mail: [danielff@dex.ufla.br](mailto:danielff@dex.ufla.br)

<sup>6</sup>Agradecimentos: a FAPEMIG, a CAPES e ao CNPq pelo apoio financeiro.

O cálculo para encontrar os arquétipos é um problema de quadrados mínimos não-linear, que pode ser resolvido por um algoritmo de otimização iterativo que converge em todos os casos, mas não necessariamente para o mínimo global. Por isso, o algoritmo deve ser iniciado várias vezes com arquétipos iniciais diferentes. A cada passo o algoritmo diminui a SQR entre a combinação linear dos arquétipos e o verdadeiro valor dos dados. O algoritmo deve parar quando a SQR for um valor suficientemente pequeno (CUTLER; BREIMAN, 1994).

Para dados multivariados  $(\mathbf{x}_i, i = 1, \dots, n)$  em que cada  $\mathbf{x}_i$  é um vetor  $p$ -dimensional  $\mathbf{x}_i = (\mathbf{x}_{1i}, \dots, \mathbf{x}_{pi})'$ , o padrão arquétipo de uma massa de dados caracteriza o problema de encontrar vetores  $p$ -dimensionais  $\mathbf{z}_1, \dots, \mathbf{z}_k$  com  $1 < k < N$ , sendo  $N$  o número de elementos na fronteira do fecho convexo dos dados (BAUCKHAGE; THURAU, 2012)  $\mathbf{z}_j = \sum_{i=1}^n \mathbf{x}_i b_{ij}$ , em que  $j = 1, \dots, k$  e os coeficientes  $b_{ij} \geq 0$  e  $\sum_{i=1}^n b_{ij} = 1$ . Assim, para uma dada escolha de arquétipos, AA minimiza  $\|\mathbf{x}_i - \sum_{j=1}^k \mathbf{z}_j a_{ji}\|^2$ .

Sabe-se que quanto maior o número de arquétipos selecionados menor é a SQR, pois menos informação é perdida, e por consequência, menor é a redução da dimensão dos dados. Então, fica a cargo do pesquisador decidir quantos arquétipos deve-se usar em um determinado conjunto de dados, desde que  $1 < k < N$ .

O esporte mais popular do planeta descobriu que intuição e improviso não são suficientes. Os times estão usando análise estatística para comprar jogadores, escalar equipes e trocar atletas durante as partidas. Como exemplo, pode citar a empresa de estatísticas de futebol (FOOTSTAT, 2014).

De acordo com o *Footstats*, constatou-se que durante a Copa de 2014, a Seleção Brasileira foi o time que mais desarmou com eficiência na competição e a Alemanha a equipe que mais tocava a bola corretamente. Uma possível explicação para estes fatos é o esquema tático utilizado por essas equipes, como por exemplo o 4-4-2, utilizado pela Seleção Brasileira na Copa do Mundo de 2014.

Nos esportes, a AA é comumente utilizada para analisar dados multivariados de atletas e redução da dimensionalidade. Os arquétipos de atletas, em geral, são os que tem desempenho muito acima ou muito abaixo da média.

## 2 Objetivos

Este trabalho tem como objetivo propor duas abordagens de uso da Análise de Arquétipos para inferir sobre a movimentação em campo de jogadores de futebol: individual e em grupos.

### 3 Metodologia

Foram simulados dados que representassem as posições ocupadas por jogadores em um campo de futebol com comprimento ( $X$ ) e largura ( $Y$ ) contínuos, limitadas entre 0 e 105 m para  $X$ , e 0 a 68 m para  $Y$  de acordo com o padrão da FIFA (2011).

Para a simulação dos dados foi utilizado uma distribuição normal bivariada, com vetor de médias e matriz de covariâncias pertinente a cada posição do jogador simulado (atacante, defesa, lateral, etc.), descritas na Tabela 1. Os valores utilizados na matriz de covariâncias foram definidos com base na área do campo que espera-se que o jogador mais atue de acordo com sua posição e característica. As covariâncias foram fixadas em 0, pois em caso contrário, os dados sorteados estariam dispostos como diagonal, o que nem sempre ocorre com dados reais. Foram sorteadas amostras independentes de tamanho 200 de uma normal bivariada truncada dentro dos limites do campo para cada jogador analisado, sendo retidos 45 pontos representando cada minuto de um tempo de um jogo.

Tabela 1: Valores utilizados para o vetor de médias e a matriz de covariâncias definidos com base na área do campo e de acordo com a posição e característica que o jogador mais atua.

Posição	Centro (X, Y)	Covariâncias
Zagueiro 1	30, 45	$\begin{pmatrix} 450 & 0 \\ 0 & 75 \end{pmatrix}$
Zagueiro 2	30, 25	$\begin{pmatrix} 450 & 0 \\ 0 & 75 \end{pmatrix}$
Lateral 1	42, 6	$\begin{pmatrix} 1000 & 0 \\ 0 & 15 \end{pmatrix}$
Lateral 2	42, 62	$\begin{pmatrix} 1000 & 0 \\ 0 & 15 \end{pmatrix}$
Meio-Campo 1	45, 20	$\begin{pmatrix} 950 & 0 \\ 0 & 75 \end{pmatrix}$
Meio-Campo 2	45, 40	$\begin{pmatrix} 950 & 0 \\ 0 & 75 \end{pmatrix}$
Meio-Campo 3	45, 60	$\begin{pmatrix} 950 & 0 \\ 0 & 75 \end{pmatrix}$
Meio-Camp 4	60, 45	$\begin{pmatrix} 950 & 0 \\ 0 & 75 \end{pmatrix}$
Atacante 1	60, 25	$\begin{pmatrix} 500 & 0 \\ 0 & 55 \end{pmatrix}$
Atacante 2	80, 35	$\begin{pmatrix} 500 & 0 \\ 0 & 55 \end{pmatrix}$

A Análise de Arquétipos foi realizada considerando duas abordagens: as posições ocupadas por um único jogador e as posições de vários jogadores simultaneamente. O goleiro não foi levado em consideração neste trabalho pois este, geralmente, atua em uma parte bem limitada e extrema do campo, porém nada impede que ele seja inserido na análise. Para gerar os dados fictícios, adotou-se uma tática semelhante a usada pela Seleção Brasileira de Futebol na Copa do Mundo de 2014, com dois zagueiros, dois laterais, quatro meio-campistas e dois atacantes.

O número de arquétipos selecionados foi definido a partir de um gráfico *ScreePlot*. Assim, após a simulação e análise dos dados foram retidos seis arquétipos para análise de um único jogador e oito arquétipos para análise de um grupo de jogadores. Estes números de arquétipos

foram definidos com o intuito de capturar melhor a forma como os dados estão distribuídos, e não só pela redução da SQR, que sugeria apenas 3 arquétipos, que é o critério comum.

## 4 Resultados e Discussão

O algoritmo apresentado a seguir descreve a abordagem sugerida neste trabalho: (i) Decidir quantos e quais jogadores serão analisados; (ii) Fazer AA com número de arquétipos igual a 6 caso para a análise de um único jogador, ou 8 caso seja analisado um grupo de jogadores; (iii) Unir os arquétipos selecionados de forma que gere um polígono; (iv) Interpretar o polígono gerado analisando as seguintes características: a área extrema percorrida, aglomeração dos pontos, perfil ofensivo/defensivo, interação entre jogadores e se há arquétipos puros.

A movimentação de todo o time de jogares em um determinado tempo de uma partida de futebol, está representada na 4(a), destacando-se que os pontos azuis foram considerados arquétipos.

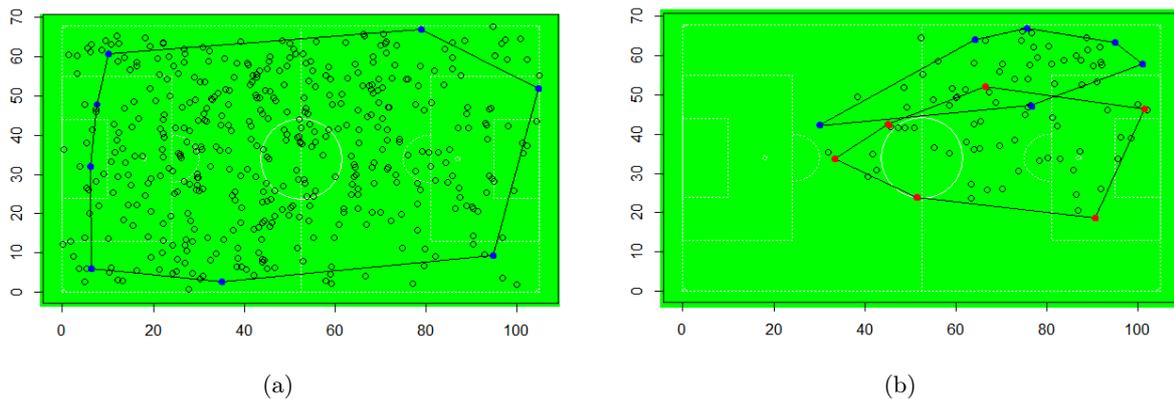


Figura 1: (a) Movimentação dos jogadores. Pontos com cores sólidas representam os arquétipos. Todos os jogadores de linha foram levados em consideração. (b) Movimentação dos dois atacantes do time. Pontos com cores sólidas representam os arquétipos. Atacante 1 em azul, atacante 2 em vermelho.

Vale ressaltar que os arquétipos podem ou não ter sido observados pelo experimento. Neste caso não houve nenhum arquétipo puro. Caso houvesse algum arquétipo puro, seria importante verificar se foi um ponto da movimentação do jogador designado para aquela área ou se foi de um jogador de outra área do campo.

Foi realizada uma nova simulação para representar a movimentação de dois jogadores do time individualmente em um determinado tempo de uma partida de futebol. A Figura 4(b) representa os resultados da análise da movimentação dos dois atacantes. Os arquétipos selecionados mostraram que ambos jogadores atuaram, mesmo que pouco, fora de seus territórios de ofício.

Nota-se também que o atacante mais avançado (arquétipos vermelhos) recuou no máximo até a intermediária defensiva, mostrando uma postura bem ofensiva, assim como o atacante mais próximo a lateral (arquétipos azuis).

Como a tática estudada, 4-4-2, apresentava apenas um jogador realmente avançado, nota-se que ele não percorreu toda a área do campo adversário, optando geralmente pelo lado superior (Figura 4(b)). O atacante mais próximo à lateral (arquétipos azuis), jogou de forma mais compactada que o atacante central, como este está mais próximo a lateral, seus pontos simulados que situarem fora dos limites do campo, serão descartados e novos pontos dentro do campo serão sorteados.

## 5 Conclusões

Com os resultados obtidos da AA de um tempo do jogo, foi possível avaliar se o grupo de jogadores analisados realmente atuam na área designada ou se eles tendem a atuar em outros setores do campo, seja para surpreender o time adversário como também para auxiliar em jogadas em outras partes do campo.

Os resultados também permitiram avaliar se o time/jogador teve um caráter mais ofensivo ou defensivo durante o tempo analisado, de acordo com o polígono construído unindo os arquétipos.

## Referências

- [1] BAUCKHAGE, C.; THURAU, C. Making Archetypal Analysis Practical. *Lecture Notes in Computer Science* v.5748, pp 272-281, 2009.
- [2] CUTLER, A.; BREIMAN, L. Archetypal analysis. *Technometrics*, v.36, pages 338-347, 1994.
- [3] FIFA - Fédération Internationale de Football Association. *Estádios de Futebol: Recomendações e requisitos técnicos*. 5ª edição, 2011. Disponível em: <[http://pt.fifa.com/mm/document/tournament/competition/01/37/17/76/p\\_sb2010\\_stadiumbook\\_ganz.pdf](http://pt.fifa.com/mm/document/tournament/competition/01/37/17/76/p_sb2010_stadiumbook_ganz.pdf)>. Acesso em: 07 jul. 2014
- [4] FOOSTATS *Estatísticas*, 2014. Disponível em: <<http://footstats.net/campeonatos/copa-do-mundo-2014/estatisticas/#eq>>. Acesso em: 28 jul. 2014
- [5] SOUZA, E. D. *Futebol paixão, produto ou identidade cultural*. 2013. 25 p. Trabalho de Conclusão de Curso (Lato Sensu em Mídia, Informação e Cultura) - Escola de Artes e Comunicações, Universidade de São Paulo, São Paulo, 2013.

# Estimativa Do Tamanho Ótimo De Parcelas Para Teste De Germinação De Sementes Da Espécie *Ceiba Speciosa*

Alexandre Antônio<sup>1</sup>

Carolina Juvelina Alves<sup>2</sup>

Douglas Silva Oliveira<sup>3</sup>

Quintiliano S. S. Nomelini<sup>4</sup>

Estudos envolvendo a determinação do tamanho ótimo de parcela são importantes no planejamento experimental e sua caracterização de forma otimizada permite a obtenção de maior precisão nesses resultados. No entanto esse tamanho não pode ser generalizado, pois deve considerar as características e especificidades de cada espécie, cultivar, local, variáveis a serem analisadas e métodos utilizados. Por esta razão, o objetivo do presente trabalho foi a utilização das metodologias: Método da Máxima Curvatura Modificado, Modelo de Regressão Linear com Platô, e Modelo de Regressão Quadrático com Platô, para a espécie *Ceiba speciosa*, em 3 lotes de variabilidades distintas para a determinação de tamanho ótimo de parcela para esta espécie. Considerou-se ensaio em branco e condições ambientais idênticas de todas as unidades experimentais básicas (UEB), com instalação dos experimentos de acordo com a metodologia proposta como sendo a melhor para teste de germinação dessa espécie segundo Nomellini (2012). Verificou-se para esta espécie que os Modelos de Regressão (Linear e Platô) reduziram os tamanhos de parcela de forma drástica e prejudicial, sendo adequado para a mesma o Método da Máxima Curvatura Modificado, e reforçando a necessidade da utilização de mais de um método para a determinação dos tamanhos de parcela.

**Palavras-chave:** Teste de Germinação, Tamanho Ótimo de Parcela, *Ceiba speciosa*

---

<sup>1</sup>FAMAT/UFU – e-mail: ac3estatistica@gmail.com

<sup>2</sup>FAMAT/UFU – e-mail: caroljuvelina@gmail.com

<sup>3</sup>FAMAT/UFU – e-mail: douglasso1988@gmail.com

<sup>4</sup>ICIAG/UFU – e-mail: dgsantana@umuarama.ufu.br

<sup>5</sup>FAMAT/UFU – e-mail: quintiliano@famat.ufu.br

# Analizando o comportamento dos turistas de Ouro Preto através de uma abordagem Espaço-Temporal

André Luís Alves Costa <sup>1</sup>

Erica Castilho Rodrigues<sup>2</sup>

Atualmente as redes sociais são responsáveis pela maior parte do tráfego na internet e são uma das principais formas de representação dos relacionamentos pessoais ou profissionais com milhões de adeptos. Elas têm se tornado cada vez mais populares e são um importante meio de intercâmbio de informações devido ao seu uso intenso, o que possibilita a reunião de um volume de dados valiosos para análises em larga escala. O *Foursquare* é uma rede social locativa, que permite ao usuário fazer um *check-in* registrando as coordenadas de sua posição naquele momento e divulgado-a para seus amigos. Esse tipo de dado é extremamente relevante para analisar o comportamento de turistas, pois permite verificar por exemplo o padrão de sua movimentação ao longo dos dias em um determinado evento; ou ainda aqueles locais que são mais visitados em uma determinada época. Neste trabalho foram analisados os dados (*check-ins*) gerados pelo *Foursquare*, durante o Festival de Inverno de Ouro Preto e Mariana do ano 2013, com o objetivo de analisar o padrão de distribuição espacial das localizações dos turistas ao longo dos dias de realização do evento. Foram aplicadas técnicas de visualização de dados e de Estatística Espacial para extrair informações relevante dos dados.

**Palavras-chave:** *Distribuição Espacial, Redes Sociais, Ouro Preto, FourSquare.*

---

<sup>1</sup>DEEST-UFOP: Universidade Federal de Ouro Preto, Ouro Preto-MG. Email: [andrealves\\_99@hotmail.com](mailto:andrealves_99@hotmail.com)

<sup>2</sup>DEEST-UFOP: Universidade Federal de Ouro Preto, Ouro Preto-MG. Email: [ericaa\\_casti@yahoo.com.br](mailto:ericaa_casti@yahoo.com.br)

# **Introdução ao estudo da curva característica de aceitação quando utilizada para discriminar planos de amostragem de aceitação de lotes de produtos**

**Laura Amaral de Oliveira**<sup>1 6</sup>

**Sandro Augusto de Faria Félix**<sup>2 6</sup>

**Thamyres Araújo de Souza**<sup>3 6</sup>

**Moema Maria Gonzaga Borges**<sup>4 6</sup>

**Andréa Cristiane dos Santos Delfino**<sup>5 6</sup>

A inspeção por amostragem é um método utilizado na indústria para aceitar ou rejeitar um lote de um determinado produto. Um dos métodos utilizados é o plano de amostragem única por atributos, neste caso, a característica da qualidade avaliada é classificada como conforme ou não conforme e o lote é sentenciado com apenas uma amostra. Uma medida importante do desempenho de um plano de amostragem de aceitação por atributos é a Curva Característica de Operação (CCO). Através da CCO é possível calcular a porcentagem de lotes que se espera aceitar, sob os vários planos de amostragem, para uma dada qualidade do processo. Neste trabalho buscou-se avaliar a CCO sob diferentes planos amostrais. As curvas foram elaboradas considerando-se a distribuição binomial com parâmetros  $n$  e  $p$ , em que  $n$  é o tamanho da amostra e  $p$  a fração de itens defeituosos. Também considerou-se o número máximo de itens defeituosos permitidos, denotado por  $c$ . Os valores de  $p$  variaram de 0,01 a 0,08, de  $c$  entre 0 a 8 e  $n$  entre 50 a 400. As CCO confeccionadas mostraram que a probabilidade de aceitar um determinado lote de produtos é menor quando a amostra é maior. Por meio dos resultados, constatou-se que a CCO pode ser utilizada como uma importante ferramenta para discriminar planos amostrais.

**PALAVRAS-CHAVE:** Amostragem por aceitação, Curva característica de operação, Distribuição Binomial.

---

<sup>1</sup>DEMEC – UFSJ: Universidade Federal de São João del Rei, São João del Rei -MG.Email: laura.oliveira4@yahoo.com.br

<sup>2</sup>DEMEC – UFSJ: Universidade Federal de São João del Rei, São João del Rei -MG.Email: sandrofelix91@yahoo.com.br

<sup>3</sup>DEMEC – UFSJ: Universidade Federal de São João del Rei, São João del Rei -MG.Email: thamyresouza@hotmail.com

<sup>4</sup>DEMEC – UFSJ: Universidade Federal de São João del Rei, São João del Rei -MG.Email: moemaborges@hotmail.com

<sup>5</sup>DEMAT – UFSJ: Universidade Federal de São João del Rei, São João del Rei -MG.Email: andrea@ufsj.edu.br

<sup>6</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## A teoria de Gauss-Markov na presença de dados adicionais

Leandro da Silva Pereira<sup>1 4</sup>

Lucas Monteiro Chaves<sup>2 4</sup>

Devanil Jaques de Souza<sup>3 4</sup>

Suponha que se tenha o modelo de Gauss-Markov  $\mathbf{Y}^1 = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , com  $\mathbf{Y}^1 \in \mathbb{R}^n$ . Suponha também que novas observações  $\mathbf{Y}^2 \in \mathbb{R}^m$  sejam obtidas. A questão é como estes dados extras podem ser incorporados ao processo de estimação. Seja  $\mathbf{Y} = \mathbf{Y}^1 + \mathbf{Y}^2$  o vetor de dados completo em  $\mathbb{R}^{n+m}$ . Quer se estender o modelo  $\Omega_1 = \text{Im}(\mathbf{X})$  em  $\mathbb{R}^n$  para  $\Omega$  em  $\mathbb{R}^{n+m}$ . Novamente a hipótese central é que o número de parâmetros não é alterado, isto é,  $\dim(\Omega_1) = \dim(\Omega)$ . Desta forma, para cada vetor  $\mathbf{v}_1 \in \Omega_1$  existe um único vetor  $\mathbf{v}_2$  tal que  $\mathbf{v}_1 + \mathbf{v}_2 \in \Omega$ . Denominando  $\Omega_2$  como o conjunto destes vetores, fica definida a transformação linear  $\mathbf{A} : \mathbb{R}^n \mapsto \mathbb{R}^{n+m}$  da forma  $\mathbf{A}(\mathbf{v}_1) = \mathbf{v}_2$  se  $\mathbf{v}_1 \in \Omega_1$  e  $\mathbf{A}(\boldsymbol{\omega}) = \mathbf{0}$  se  $\boldsymbol{\omega} \in \Omega_1^\perp$ .

A idéia para se estimar  $\boldsymbol{\mu}$  é projetar ortogonalmente o vetor de dados completos  $\mathbf{Y} = \mathbf{Y}^1 + \mathbf{Y}^2$  no subespaço  $\Omega$  e depois projetar novamente no subespaço  $\Omega_1$  do modelo original  $\mathbf{X}\hat{\boldsymbol{\mu}} = P_{\Omega_1}P_\Omega(\mathbf{Y}^1 + \mathbf{Y}^2)$ .

**Palavras-chave:** *Subespaços vetoriais, Dados adicionais, Matrizes de projeção.*

---

<sup>1</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [leandropereira@dex.ufla.br](mailto:leandropereira@dex.ufla.br)

<sup>2</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [lucas@dex.ufla.br](mailto:lucas@dex.ufla.br)

<sup>3</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [devanil@dex.ufla.br](mailto:devanil@dex.ufla.br)

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

# Gráficos de controle utilizando o *software R*: aplicação em uma indústria agroquímica de Uberlândia-MG

Maxwell Nunes do Carmo<sup>1 3</sup>

Orientadora Dra. Aurélia Aparecida de Araújo Rodrigues<sup>2 3</sup>

## Resumo

*O controle estatístico de processos é uma poderosa ferramenta no auxílio da qualidade industrial, no auxílio contra perdas, falhas, multas e prejuízos provenientes da falta de controle de um determinado processo. Neste contexto pode-se usar os gráficos de controle que podem detectar e identificar as causas responsáveis por variações que afetam o comportamento do processo de maneira não prevista ou indesejada. O objetivo deste trabalho é utilizar o Software estatístico R na construção do gráfico de controle para monitorar o peso de caminhões tanque que transportam aditivos para fertilizantes da empresa Alfa e verificar se o processo está estatisticamente em controle.*

*Palavras chave: Controle Estatístico de Processos, Software R, gráfico de controle.*

## 1. Introdução

A cada dia as empresas tornam seus processos e produtos mais enxutos, reduzindo os custos diretos com os produtos, aumentando assim sua competitividade no mercado. Neste cenário pode-se então observar a importância do estudo do Controle Estatístico de Processos (CEP), para entender o fluxo existente nas indústrias bem como as possíveis melhorias para a redução de custos e eliminação de desperdícios e processos que não agregam valor ao produto final.

Assim, o CEP é um poderoso conjunto de ferramentas utilizado na obtenção de estabilidade do processo e na melhoria de sua capacidade através da redução da variabilidade (DINIZ *et al.*, 2013).

O presente trabalho foi desenvolvido na empresa Alfa<sup>4</sup> do setor Agroquímico que oferece soluções com produtos destinados ao Agrobusiness<sup>5</sup>. Com apenas quatro anos de existência a Empresa Alfa se tornou a maior empresa do país em fabricação e distribuição de aditivos para beneficiamento de fertilizantes granulados, atendendo clientes no Brasil e outros países da América do Sul, África e Europa.

---

<sup>1</sup> FAMAT-UFU: Universidade Federal de Uberlândia, Uberlândia – MG. Email: maxwell.carmo@yahoo.com.br

<sup>2</sup> FAMAT-UFU: Universidade Federal de Uberlândia, Uberlândia – MG. Email: aurelia@famat.ufu.br

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

<sup>4</sup> Por questões éticas o nome verdadeiro da empresa não será divulgado neste trabalho. Por este motivo o nome foi substituído por Alfa.

<sup>5</sup> Segundo o Portal do Agronegócio, Agrobusiness é definido como a soma das operações de produção e distribuição de suprimentos agrícolas, processamentos e distribuição dos produtos agrícolas e itens produzidos a partir deles.

## 2. Objetivo

O objetivo geral deste trabalho é utilizar o *Software* estatístico *R* para a construção de gráficos de controle para monitorar o peso carregamento de caminhões tanque que transportam produto químico.

## 3. Metodologia

A ferramenta utilizada para monitorar os processos de produção é o gráfico de controle. Para sua operação, o processo é monitorado da seguinte maneira: periodicamente, retira-se uma amostra de certo número de itens do processo, calcula-se uma ou mais estatísticas amostrais (como medidas individuais, média, amplitude ou desvio-padrão) e registram-se seus valores no(s) gráfico(s) de controle.

O gráfico de controle consiste de limite inferior de controle (*LIC*), linha média (*LM*) e limite superior de controle (*LSC*). A variação devida a causas especiais é sinalizada por pontos acima do *LSC* ou abaixo *LIC*. As causas especiais de variação devem ser, de modo geral, localizadas e eliminadas. E, além disso, devem-se adotar medidas para evitar sua reincidência. Se nenhuma causa especial é sinalizada pelo gráfico e somente causas comuns estão agindo no processo, diz-se que o processo está *em controle*. Caso contrário, isto é, se alguma causa especial estiver atuando no processo, tem-se, então, um processo *fora de controle*. Mais detalhes sobre gráficos de controle podem ser obtidos em COSTA *et. al.* (2012) e MONTGOMERY (2004)

Neste trabalho, como o carregamento de caminhões tanque possui taxa de produção baixa, será aplicado o gráfico de controle para medidas individuais com os seguintes limites de controle:

$$LSC = \bar{x} + 3(\overline{AM}/d_2)$$

$$LM = \bar{x}$$

$$LIC = \bar{x} - 3(\overline{AM}/d_2)$$

em que,  $MR_i = |x_i - x_{i-1}|$  e  $d_2 = 1,128$ .

Os dados deste trabalho foram coletados *in loco*, ou seja, foram coletados no local onde foram gerados. As coletas não seguiram nenhum intervalo determinado de tempo para serem gerados, pois os carregamentos de caminhões ocorriam diariamente, sendo de 0 a 5 carregamentos por dia, em horários aleatórios.

Quando o caminhão chega na empresa Alfa, ele é posicionado em seu local devido para fazer o carregamento. Quando finalizado o posicionamento um operador posiciona a lança de carregamento na parte superior do tanque, para então acionar uma bomba centrífuga que fará o transporte do produto do tanque de estocagem até o tanque do caminhão. Todo este fluxo de produto é controlado por um medidor volumétrico que mede a quantidade de volume que foi depositado dentro do caminhão.

Ao finalizar o carregamento, o caminhão passa por uma balança rodoviária permitindo a identificação do peso final. A diferença entre o peso final e o peso do caminhão vazio é o peso do produto dentro do caminhão. Os dados apresentados na tabela 1 foram obtidos no mês de Outubro de 2013.

**Tabela 1:** Peso do carregamento de produto químico, em toneladas

Obs	Peso								
01	29,64	11	31,28	21	29,41	31	31,64	41	31,98
02	29,65	12	31,00	22	31,70	32	31,94	42	29,61
03	31,34	13	31,29	23	31,78	33	30,01	43	31,46
04	31,61	14	31,92	24	30,04	34	29,48	44	31,53
05	29,31	15	31,47	25	31,99	35	31,72	45	29,59
06	29,46	16	31,62	26	31,25	36	29,35	46	31,88
07	29,41	17	31,09	27	31,05	37	31,23	47	29,65
08	31,29	18	29,68	28	31,65	38	29,60	48	31,29
09	29,46	19	31,58	29	29,65	39	31,10	49	29,53
10	30,82	20	30,77	30	29,60	40	31,24	50	31,32

O *Software R* dispõe de três pacotes específicos para aplicações de CEP: o qcc (Quality Control Charts), o spc (Statistical Process Control) e o qAnalyst. Aqui será empregada a função qcc para a construção do gráfico de controle para medidas individuais.

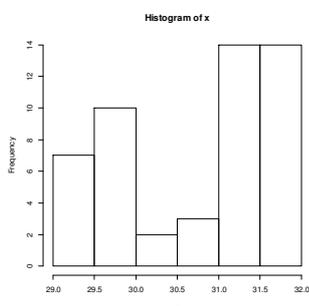
#### 4. Resultados e Discussões

Para a análise exploratória dos dados, utilizou-se o *software R* para gerar as medidas descritivas, o histograma, o box-plot.

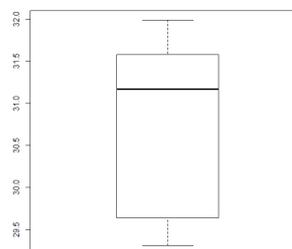
**Tabela 2:** Medidas estatísticas, em toneladas

Média	30.74	1° Quartil	29.63
Mediana	31.16	2° Quartil	31.16
Moda	29.65	3° Quartil	31.58
Variância	0.91	Coefficiente de assimetria (As)	-2,39
Valor Máximo	31.99	Coefficiente de Variação	3,11%
Valor Mínimo	29.31		

De acordo com a análise exploratória dos dados apresentada na tabela 1, destaca-se que o peso médio do carregamento é 30,74 toneladas e o coeficiente de variação foi de 3,11%.



**Figura 1:** Histograma



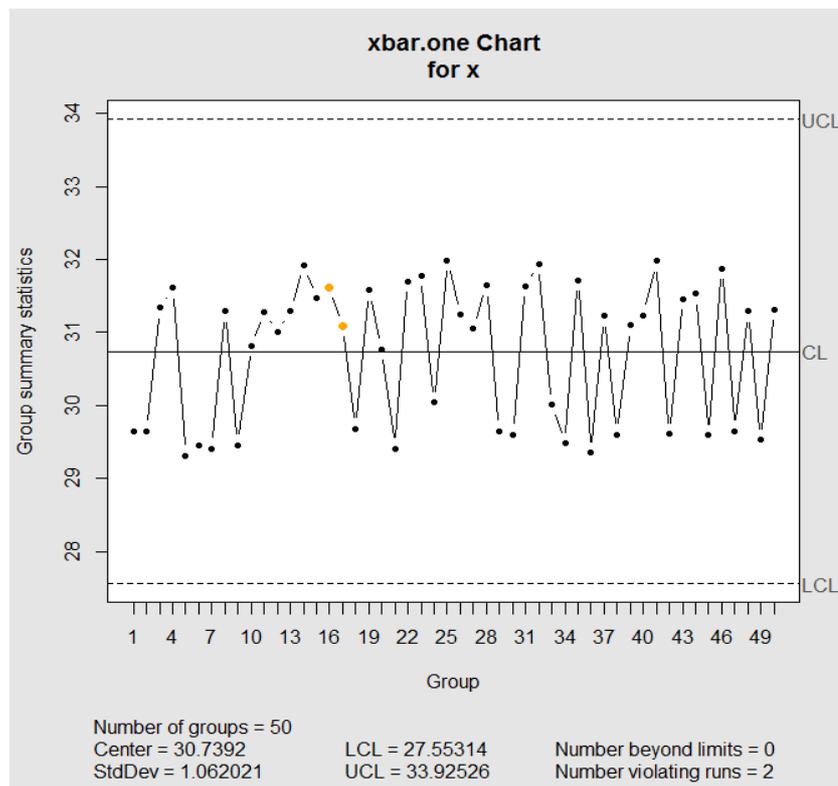
**Figura 2:** Box plot

No entanto, a mediana também é adequada para representar este conjunto de dados, pois a distribuição é assimétrica à esquerda ( $As = -2,39$ ). No box-plot da Figura 2, percebe-se que não há valores discrepantes no conjunto de dados.

De acordo com o teste de Shapiro-Wilk, os dados não apresentam distribuição normal, pois o  $p$ -valor deste teste é 0,0091.

Para gerar o gráfico de controle para medidas individuais da Figura 3, foram implementados os seguintes comandos no *software R*:

```
require(qcc)
qcc(data=x,type="xbar.one")
```



**Figura 3:** Gráfico de controle para medidas individuais

Através do gráfico é possível verificar que não houve valores fora dos limites de controle. No entanto, o gráfico apresentou sinal de alerta nos pontos 16 e 17. Ou seja, há uma sequência de sete pontos acima da linha média. Após investigação, verificou-se que não houve causas especiais atuando no processo. Então, para monitoramento do processo, pode-se utilizar os limites LIC = 27,55 e LSC = 33,93, os quais serão comparados com as futuras observações.

Os resultados apresentados referem-se a um estudo preliminar do processo de enchimento de caminhões tanque com produtos químicos. Como continuidade deste trabalho, pretende-se propor limites mais estreitos ( $k < 3$ ), empregar o gráfico de controle para amplitude móvel para monitorar a variabilidade, obter as medidas de desempenho do gráfico de controle e estudar os efeitos da violação da suposição de normalidade.

## 5. Conclusão

Neste trabalho, foi aplicado o gráfico de controle para medidas individuais para monitorar o processo de carregamento de caminhões tanque que transportam produtos químicos produzido pela empresa Alfa. Após investigação do sinal de descontrole do gráfico de controle para medidas individuais verificou-se que processo não havia sofrido influências de causas especiais. Assim sendo os limites obtidos podem ser empregados para monitorar as observações futuras.

O processo de carregamento de caminhões tanque que transportam produtos químicos produzido pela empresa Alfa deve ser monitorado continuamente. Recomenda-se rever periodicamente os limites, mesmo que o processo permaneça estável.

A implementação de técnicas de controle estatístico de processo, em qualquer ambiente, pode ser feita por um funcionário que possua conhecimentos de estatística básica e seja adequadamente treinado para implementar gráficos de controle.

## Referências

COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. C. R. *Controle Estatístico de Qualidade*. São Paulo-SP : Atlas, 2012.

DINIZ, C.; ET. AL. *Controle Estatístico de Processos: Uma abordagem prática para cursos de Engenharia e Administração*. Rio de Janeiro : LTC, 2013.

MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. 5. ed. New York: JohnWiley, 2004.

# Estudo de divergência genética para características de qualidade de carne entre diferentes linhagens de codorna

**Thaiza da Silva Campideli<sup>1 2</sup>**

**Talita Andrade Ferreira<sup>1 2</sup>**

**Luiza Rodrigues Alves Abreu<sup>1 2</sup>**

**Aldrin Vieira Pires<sup>1 2</sup>**

**Adalfredo Rocha Lobo Júnior<sup>1 2</sup>**

Os programas de melhoramento possui como ferramenta a análise de agrupamentos por meio de uma metodologia multivariada. Portanto, objetivou-se verificar, por meio de análise multivariada, a divergência genética entre linhagens de codornas para características de qualidade de carne. Um total de 160 codornas, provenientes de sete linhagens de codornas de corte (EV1, EV2, UFV1, UFV2, UFV3, LF1 e LF2) e uma de postura (POST) foram utilizadas. As codornas foram abatidas aos 35 dias de idade e coletadas amostras de carnes para realizar as análises: determinação de pH, capacidade de retenção de água, perda de peso por cozimento, força de cisalhamento e coloração ( $L^*$ ,  $a^*$ ,  $b^*$ , C e H). As análises estatísticas foram realizadas utilizando o “proc glm” do SAS. Para testar contrastes entre as médias foram calculadas as diferenças mínimas significativas ao nível probabilidade de 5% pelo teste de Roy. Na análise de agrupamento utilizou-se o método de otimização de Tocher, adotando-se a distância D2 de Mahalanobis como medida de dispersão. Não houve diferença ( $p > 0,05$ ) entre as linhagens avaliadas para as características de qualidade da carne, usando-se o teste de Roy. A análise de agrupamento pelo método de otimização de Tocher resultou em dois grupamentos genéticos distintos. A linhagem POST apresentou maior distanciamento de produção comparado às demais linhagens. Conclui-se que todas as linhagens de codornas tem o mesmo potencial para as propriedades relacionadas à qualidade da carne.

**Palavras-chave:** *Multivariada, teste de Roy, otimização de Tocher.*

---

<sup>1</sup>DZO – UFVJM: Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina-MG. Email: [thaizazootecnista@gmail.com](mailto:thaizazootecnista@gmail.com)

<sup>2</sup>Agradecimento à FAPEMIG e CNPq pelo apoio financeiro.

# Modelos da Teoria de Resposta ao Item assimétricos de grupos múltiplos para respostas politômicas nominais e ordinais sob um enfoque bayesiano

Eduardo Vargas Ferreira<sup>1 3</sup>

Caio Lucidius Naberezny Azevedo<sup>2 3</sup>

No presente trabalho propõem-se novos modelos da Teoria de Resposta ao Item para respostas politômicas nominais e ordinais, via dados aumentados, para grupos múltiplos. Para a modelagem das distribuições dos traços latentes considera-se distribuições normais assimétricas centradas. Com relação aos métodos de estimação, desenvolveu-se procedimentos bayesianos através de algoritmos de Monte Carlo via cadeias de Markov (MCMC), utilizando o algoritmo de Gibbs, com a verossimilhança aumentada (dados aumentados) e *Metropolis-Hastings*, considerando a verossimilhança original. As implementações computacionais foram escritas em linguagem C++, integradas ao ambiente R, viabilizando rotinas gratuitas, de código aberto e alta velocidade no processamento, essenciais à difusão de tais metodologias. Para a seleção de modelos, utilizou-se o critério de informação *deviance* (DIC), os valores esperados do critério de informação de Akaike (EAIC) e o critério de informação bayesiano (EBIC). Em relação à verificação da qualidade do ajuste de modelos, explorou-se a checagem preditiva a posteriori. Estudos de simulação indicam que os modelos e métodos de estimação produzem resultados bastante satisfatórios, com superioridade dos modelos assimétricos com relação ao simétricos. A análise de um conjunto de dados reais, referente à primeira fase do vestibular da UNICAMP de 2013, ilustra o potencial da tríade: modelagem, métodos de estimação e ferramentas de diagnósticos, desenvolvida neste trabalho.

**Palavras-chave:** *Teoria da resposta ao item, modelos politômicos, inferência bayesiana, distribuição normal assimétrica centrada, algoritmos MCMC, dados aumentados.*

---

<sup>1</sup>IMECC-UNICAMP: Universidade Estadual de Campinas, Campinas -SP. Email: [eduardozenus@yahoo.com.br](mailto:eduardozenus@yahoo.com.br)

<sup>2</sup>IMECC-UNICAMP: Universidade Estadual de Campinas, Campinas -SP. Email: [cnaber@ime.unicamp.br](mailto:cnaber@ime.unicamp.br)

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

# Classificação de curvas de maturação de cana-de-açúcar utilizando técnicas multivariadas

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

Luiz Alexandre Peternelli<sup>2 3</sup>

As curvas de maturação são utilizadas como ferramenta para planejamento estratégico visando aumentar o rendimento ao longo dos meses da safra de cana-de-açúcar, de modo a obter o máximo de lucro possível. Em geral, essas curvas apresentam formato de um polinômio de segundo grau que podem ser utilizadas para comparar os clones, agrupando-os em classes de maturação precoce, média ou tardia. Adicionalmente esses clones, já em fase final de experimentação dentro do programa de melhoramento da cana, são comparados com padrões (variedades comerciais) de classe de maturação conhecida. O objetivo deste trabalho é avaliar a aplicabilidade do método apresentado por Peternelli et al. (2013) para agrupar os clones quanto a sua curva de maturação. Foram utilizadas as curvas de maturação de clones da fase final de experimentação juntamente com alguns padrões comerciais do Programa de Melhoramento de Cana da Universidade Federal de Viçosa (PMGCA-UFV). Estimativas dos coeficientes de regressão de cada modelo polinomial foram organizadas numa matriz a ser usada nas análises seguintes. Na análise de agrupamento dos clones foi usada a distância euclidiana seguido do método de Ward para formação dos três grupos acima definidos. A classificação final foi comparada com aquelas previamente definidas pelos melhoristas. Alguns resultados preliminares de avaliação do uso dessa proposta para a classificação das curvas de maturação de clones serão apresentados.

**Palavras-chave:** seleção de cana-de-açúcar; análise multivariada; modelo polinomial.

---

<sup>1</sup>DET - CCE – UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [jaqueline.goncalves@ufv.br](mailto:jaqueline.goncalves@ufv.br)

<sup>2</sup>DET - CCE – UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [peternelli@ufv.br](mailto:peternelli@ufv.br)

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

# Aplicações de técnicas biométricas no estudo do comprimento de hipocótilo de cultivares de soja

Ronaldo Machado Junior<sup>1 2</sup>

Éder Matsuo<sup>1 2</sup>

Guilherme Ferreira Alves<sup>1</sup>

Luiz Henrique de Oliveira Dianin<sup>1</sup>

Pesquisas envolvendo descritores adicionais tem sido de grande importância para o melhoramento genético de soja. Objetivou-se neste trabalho, aplicar técnicas biométricas no estudo do comprimento de hipocótilo de cultivares de soja. O trabalho foi desenvolvido em casa de vegetação da Universidade Federal de Viçosa – *Campus*Rio Paranaíba, onde foram analisados sete cultivares no delineamento inteiramente casualizado com quatro repetições, sendo cada repetição representada pela média de 10 plantas. Com auxílio de paquímetro digital foi mensurado o comprimento do hipocótilo quando as plantas estavam no estágio de desenvolvimento V2. Os dados obtidos foram submetidos ao teste F na análise de variância que indicou variabilidade genética entre cultivares. Em seguida foram estimados os parâmetros genéticos (coeficiente de variação experimental (C<sub>Ve</sub>) e genético (C<sub>Vg</sub>), relação C<sub>Vg</sub>/C<sub>Ve</sub> e coeficiente de variação genotípico (H<sup>2</sup>)) e realizado a análise de repetibilidade. As estimativas de parâmetros genéticos evidenciaram situação favorável à identificação de genótipos superiores e com 15 medições obteve-se 85% de confiabilidade pelos métodos Anova e Análise estrutura e 90% para Componente Covariância e Componente Correlação.

**Palavras chave:** *Glycinemax*, Descritores adicionais, Melhoramento genético.

---

1 Universidade Federal de Viçosa – *Campus* Rio Paranaíba. Emails: [ronaldo.machado@ufv.br](mailto:ronaldo.machado@ufv.br), [edermatsuo@ufv.br](mailto:edermatsuo@ufv.br), [guilherme.f.alves@ufv.br](mailto:guilherme.f.alves@ufv.br) e [luiz.dianin@ufv.br](mailto:luiz.dianin@ufv.br)

2 Agradecimento à FAPEMIG pelo apoio financeiro.

# Comparação entre as abordagens bayesiana e frequentista na avaliação de curvas de crescimento de suínos genotipados para o gene halotano

Leandro Roberto de Macedo<sup>1 4</sup>

Paulo Roberto Cecon<sup>1 4</sup>

Fabyano Fonseca e Silva<sup>2 4</sup>

Jussara Aparecida dos Santos<sup>3 4</sup>

Para avaliar a influência do gene halotano sobre as características de crescimento em suínos, bem como sua interação com o sexo do animal, foram propostas duas diferentes abordagens: Frequentista e Bayesiana, onde foi estudado um total de 344 suínos F2(comercial x Piau), machos castrados e fêmeas, pesados nos seguintes dias: ao nascimento, aos 21, 42, 63, 77, 105 e 150 (abate) dias. A modelagem hierárquica Bayesiana consistiu no ajuste do modelo de regressão não-linear Logístico, que foi aplicado aos dados de peso-idade de suínos, por meio da qual é possível contemplar simultaneamente a estimação dos parâmetros do modelo não-linear e estimação do efeito do gene sobre as estimativas destes parâmetros, e também testar a significância do efeito do gene por meio da confecção de intervalos HPD. As amostras da distribuição marginal *a posteriori* para as diferenças entre as estimativas dos parâmetros do modelo indicaram que o peso dos suínos machos à idade adulta com genótipo heterozigoto (Nn) é superior ao dos homozigotos (NN). Por outro lado, a abordagem Frequentista não encontrou diferenças significativas em relação ao efeito do gene bem como sua interação com o sexo. A abordagem Bayesiana mostrou-se mais adequada por estimar parâmetros mais condizentes com a realidade, além de uma maior facilidade de convergência.

**Palavras-chave:** *regressão não-linear, abordagem bayesiana, abordagem frequentista.*

---

<sup>1</sup> DET – UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [leandro.macedo@ufv.br](mailto:leandro.macedo@ufv.br); [cecon@ufv.br](mailto:cecon@ufv.br)

<sup>2</sup> DZO – UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [fabyanofonseca@ufv.br](mailto:fabyanofonseca@ufv.br)

<sup>3</sup> DFT – UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [jussara.santos@ufv.br](mailto:jussara.santos@ufv.br)

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

# Abordagem Bayesiana da forma linearizada da Isoterma de Freundlich

Gustavo Borges Vieira<sup>1 3</sup>

Denismar Alves Nogueira<sup>2 3</sup>

Luiz Alberto Beijo<sup>3 3</sup>

Objetivou-se avaliar o uso da Inferência Bayesiana no ajuste da forma linearizada da Isoterma de Freundlich [ $\log(Y) = \log(K) + b \cdot \log(C)$ ] a uma amostra de tamanho cinco, sobre dados de adsorção de  $\text{Cd}^{+2}$ , extraídos de Canellas et. al (1990). Na isoterma de Freundlich  $K$  é o parâmetro de afinidade,  $C$  é a concentração,  $b$  é o parâmetro do grau de heterogeneidade da superfície do adsorvente e  $Y$ , a quantidade adsorvida. A abordagem Bayesiana tem apresentado importantes resultados pelo fato de permitir a incorporação, no processo de inferência, de informações *a priori* sobre os parâmetros de um modelo. Foram pesquisadas na literatura informações sobre a média e variância dos valores estimados para  $K$  e  $b$  da Isoterma de Freundlich. Linhares et al. (2009) afirmaram que a média dos valores dos logaritmos naturais de  $K$  é de  $2,749 \mu\text{g} \cdot \text{g}^{-1}$ , com desvio padrão de  $0,907 \mu\text{g} \cdot \text{g}^{-1}$ , e para  $b$  a média de  $1,341 \mu\text{g} \cdot \text{g}^{-1}$  e o desvio padrão de  $0,230 \mu\text{g} \cdot \text{g}^{-1}$ . Após o cálculo da distribuição *a posteriori* assumindo normalidade para a verossimilhança e uma priori normal-gama para  $K, b$  e  $\tau$ , estimou-se a equação ajustada em  $\hat{Y} = 3,39 + 0,866C$ , sendo o intervalo HPD de 95% para  $\ln K$  de  $[2,739; 4,035]$  e para  $b$  de  $[0,660; 1,059]$ . Esta abordagem mostrou-se eficiente para estimar parâmetros da Isoterma de Freundlich, fornecendo resultados precisos, mesmo para amostra relativamente pequena.

**Palavras-chave:** Adsorção, Íons de Cádmiio, Priori Informativa .

---

<sup>1</sup>ICEx - UNIFAL-MG: Universidade Federal de Alfenas, Alfenas-MG. Email: [gborgesvieira@gmail.com](mailto:gborgesvieira@gmail.com)

<sup>2</sup>ICEx - UNIFAL-MG: Universidade Federal de Alfenas, Alfenas-MG. Email: [denisnog@yahoo.com.br](mailto:denisnog@yahoo.com.br)

<sup>3</sup>ICEx - UNIFAL-MG: Universidade Federal de Alfenas, Alfenas-MG. Email: [prof.beijo@gmail.com](mailto:prof.beijo@gmail.com)

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

# Estudo Comparativo de Incidência de Temperaturas Extremas no Município de Machado-MG

Eleanderson Campos Eugênio Filho<sup>1 3</sup>

Deive Ciro de Oliveira<sup>2 3</sup>

O café é um dos produtos agrícolas mais importantes para o Brasil. O Sul de Minas é referência em produção e o município de Machado contribui de forma relevante para isto. O café é sensível à temperaturas baixas, que podem afetar no desenvolvimento e produção da lavoura. Visto isso, este trabalho tem por objetivo ajustar e comparar dois processos de contagem relativos à incidência de geadas no município de Machado nos períodos de 1962-1984 e 1991-2013.

Os dados de temperatura mínima diária foram obtidos do Instituto Nacional de Meteorologia (INMET). O evento considerado foi a geada, definida por temperaturas mínimas menores que  $2,5^{\circ}C$ . Para a modelagem do evento, foi adotado o processo de contagem de Poisson, onde a intensidade de ocorrência do evento  $\lambda(d)$  é uma função dos dias do ano ( $d \in [0, 365]$ ). A estimação e comparação das funções intensidade foi feita no software *R*. As estimativas foram obtidas pelo método de máxima verossimilhança e a comparação se deu via teste de hipóteses, admitindo normalidade para os estimadores dos parâmetros.

As estimativas para  $\lambda(d) = e^{\beta_0 d^2 + \beta_1 d + \beta_2}$  foram:  $\hat{\lambda}_1(d) = e^{-0,0006d^2 + 0,2515d - 29,6023}$  para a janela de 1962-1984 e  $\hat{\lambda}_2(d) = e^{-0,0017d^2 + 0,6435d - 65,2950}$  para 1991-2013. Considerando a hipótese nula de igualdade de  $\beta_0, \beta_1, \beta_2$  para as duas funções estudadas, os p-valores foram de 0,2015; 0,2119 e 0,2297 respectivamente. Os resultados não evidenciaram mudanças na incidência de geadas no município de Machado nos períodos analisados.

**Palavras-chave:** *Eventos extremos, Café, Processo de Poisson.*

---

<sup>1</sup>discente ICESA - UNIFAL-MG. Email: [eleandersoncampos@gmail.com](mailto:eleandersoncampos@gmail.com)

<sup>2</sup>docente ICESA - UNIFAL-MG. Email: [deive.oliveira@unifal-mg.edu.br](mailto:deive.oliveira@unifal-mg.edu.br)

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

# Uma aplicação dos modelos de fragilidade compartilhada no melhoramento de suínos

Vinicius Silva dos Santos<sup>1 5</sup>

Sebastião Martins Filho<sup>2 5</sup>

Fabyano Fonseca e Silva<sup>3 5</sup>

Marcos Deon Vilela de Resende<sup>4 5</sup>

Em programas de melhoramento genético animal há interesse em selecionar indivíduos que apresentem menor tempo possível para atingir um peso desejável. Um dos modelos mais utilizados em análise de sobrevivência multivariada é o modelo de fragilidade, em que um efeito aleatório denominado fragilidade é introduzido no modelo de regressão de Cox. Assim, o objetivo deste trabalho foi comparar os modelos de fragilidade compartilhada com distribuições Gama e Log-normal em um conjunto de dados reais de 345 suínos cedidos pelo departamento de zootecnia da UFV. A variável resposta foi o tempo (em dias) que um animal levou para ganhar o peso ideal do nascimento até o abate e as covariáveis utilizadas foram sexo, lote de manejo e presença do gene halotano. A comparação dos modelos foi baseada nas estimativas dos efeitos fixos e intervalos de confiança, além da análise gráfica em ambos os modelos de fragilidade e no modelo usual de Cox. Foram comparadas também as medidas de ajuste: Teste de Wald, variância do efeito aleatório, *Rsquare*, AIC e BIC. Os resultados demonstraram que a inclusão da fragilidade permitiu uma melhora na qualidade de ajuste em relação ao modelo de Cox clássico. Apesar dos critérios utilizados indicarem como melhor o modelo de fragilidade log-normal, nota-se que ambos os modelos de fragilidade apontaram os mesmos grupos com risco diferenciado e estimativas de efeitos fixos e intervalos de confiança semelhantes.

**Palavras-chave:** *efeito aleatório, censura, seleção.*

---

<sup>1</sup>DET - Universidade Federal de Viçosa. Email: [vinicius.s.santos@ufv.br](mailto:vinicius.s.santos@ufv.br)

<sup>2</sup>DET - Universidade Federal de Viçosa. Email: [martinsfilho@ufv.br](mailto:martinsfilho@ufv.br)

<sup>3</sup>DET - Universidade Federal de Viçosa. Email: [fabyanofonseca@ufv.br](mailto:fabyanofonseca@ufv.br)

<sup>4</sup>DET - Universidade Federal de Viçosa. Email: [marcos.deon@gmail.com](mailto:marcos.deon@gmail.com)

<sup>5</sup>Agradecimento à CAPES, CNPq e FAPEMIG pelo apoio financeiro.

# Análise espacial do processo de restauração na região da usina hidrelétrica de Camargos, Itutinga-MG

Gilberto Rodrigues Liska <sup>1 6</sup>

Luciana Maria de Souza <sup>2 6</sup>

Marcelo Ângelo Cirillo <sup>3 6</sup>

Guido Gustavo Humada-Gonzalez <sup>4 6</sup>

Estevan Rodrigues Liska <sup>5 6</sup>

A degradação dos ambientes naturais tem sido um grande desafio a ser enfrentado, quando um ecossistema é degradado, danificado, transformado ou totalmente destruído como resultado direto ou indireto das atividades humanas torna-se necessário aplicar técnicas que visem à restauração ecológica. Para analisar situações do tipo, indicadores ecológicos são utilizados, entre eles, o fechamento do dossel, representado pela variável índice de fechamento do dossel (IFD), tem se apresentado como um bom indicador, uma vez que controla a quantidade, a qualidade e a distribuição temporal e espacial da luz. O presente estudo foi conduzido no entorno da Usina Hidrelétrica de Camargos, município de Itutinga-MG. As áreas estudadas passaram por um processo de degradação no período de construção da barragem a partir 1950 e após cerca de 30 anos iniciou-se a recuperação dessas áreas. Diante do exposto, objetivou-se analisar a variável IFD, a fim de viabilizar informações sobre sua distribuição espacial, utilizando recursos da Geoestatística. Os resultados apontam o modelo Exponencial como mais adequado. A estimativa para o alcance foi de 877,97 metros, ou seja, até essa distância existem evidências estatísticas de dependência espacial sobre a distribuição do IFD. A Krigagem Ordinária foi utilizada para efetuar as predições de IFD e os resultados mostram a ocorrência de várias regiões com valores superiores a 80% de IFD, o que indica que o processo de restauração da região tem sido alcançado com sucesso.

**Palavras-chave:** *Índice de fechamento do dossel, Krigagem ordinária, Indicador ecológico, Restauração ecológica.*

---

<sup>1</sup>DEX - Universidade Federal de Lavras. Email: [gilbertoliska@hotmail.com](mailto:gilbertoliska@hotmail.com)

<sup>2</sup>DEX - Universidade Federal de Lavras. Email: [vilasboaslu@yahoo.com.br](mailto:vilasboaslu@yahoo.com.br)

<sup>3</sup>DEX - Universidade Federal de Lavras. Email: [marcelocirillo@gmail.com](mailto:marcelocirillo@gmail.com)

<sup>4</sup>DEX - Universidade Federal de Lavras. Email: [gustavohumad@hotmail.com](mailto:gustavohumad@hotmail.com)

<sup>5</sup>ICN - Universidade Federal de Alfenas. Email: [estevanliska@gmail.com](mailto:estevanliska@gmail.com)

<sup>6</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

## **Análise de co-integração e causalidade das séries: produção e preço do petróleo e consumo aparente de derivados do petróleo**

**Tarcius Camilo da Silva**<sup>1 3</sup>

**Luciene Resende Gonçalves**<sup>2 3</sup>

O petróleo é uma das matérias-primas de maior significância tanto para o balanço comercial dos países quanto para balanço contábil das empresas. Com isso é importante buscar uma melhor tomada de decisão, sendo imprescindível obter informações sobre produção e preço do petróleo e consumo aparente de derivados do petróleo. Então neste trabalho, buscou-se identificar a relação de causalidade entre produção e preço do petróleo e consumo aparente de derivados do petróleo no Brasil utilizando o teste de co-integração e causalidade de Engle-Granger. O roteiro metodológico consistiu no levantamento de dados históricos e quantitativos referentes ao preço e produção de petróleo e consumo aparente de derivados do petróleo. Utilizou-se o programa *Gretl* para geração de gráficos, realização do teste de raiz unitária, análise de co-integração entre as variáveis, verificação de causalidade e estruturação do modelo VAR para obtenção de previsões. Como resultado, identificou-se através de análise gráfica e também por meio do teste de Dickey-Fuller que todas as séries analisadas são não estacionárias. Realizou-se, então, o teste de co-integração. Os resultados do teste de causalidade indicaram que o preço do petróleo e o consumo aparente de derivados do petróleo apresentaram bi-causalidade. Dessa forma ajustou-se um modelo VAR e obteve previsões no período de setembro de 2013 a dezembro de 2013 para estas séries. Verificou-se a existência de causalidade unilateral entre as demais variáveis.

**Palavras-chave:** *Causalidade de Granger, modelo VAR, teste de raiz unitária.*

---

<sup>1</sup> UNIFAL: Universidade Federal de Alfenas – *Campus Varginha*, Varginha-MG. Email: [tarcius\\_camilo@hotmail.com](mailto:tarcius_camilo@hotmail.com)

<sup>2</sup> ICESA – UNIFAL: Universidade Federal de Alfenas – *Campus Varginha*, Varginha-MG. Email: [Luciene.goncalves@unifal-mg.edu.br](mailto:Luciene.goncalves@unifal-mg.edu.br)

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

# Otimização das estimativas dos parâmetros de limiar em testes sensoriais

Paulo César Moraes Ribeiro <sup>1</sup>

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

Charles Shalimar Filipe da Silva<sup>3</sup>

No planejamento de experimentos um papel importante é desempenhado pela matriz de informação de Fisher. Para que se defina por um experimento ser melhor ou pior que outro, calcula-se resumos de tal matriz. Em modelos lineares ( $E[y] = X\theta$ ), esta matriz depende apenas da matriz de delineamento  $X$ . Em modelos não lineares, incluindo aqui os lineares generalizados e mistos, temos que  $G(E[y|\mu]) = X\beta + Z\mu$ . Assim a Informação não depende apenas de  $(X|Z)$  que é a matriz de delineamento, mas da distribuição de  $\mu$  e da função  $G$  que se aplica (como a normas ou t de Student acumuladas). Este trabalho visa montar um programa que, dado um valor dos parâmetros e distribuições associadas a  $\theta' = (\beta', \mu')'$ , calcula-se um critério de informação para dizer se uma disposição de  $X|Z$  é melhor ou pior que outra. A aplicação prática disso é:  $Z$  é a parte da matriz do delineamento referente a provadores, que não muda (é o que for disponível em termos de provadores...), mas  $X$  é a parte referente aos tratamentos, no caso, três linhas de regressão com interceptos comuns. Assim é possível definir valores melhores para  $X$  em um novo experimento de forma a estimar melhor o limiar de sensibilidade. De modo que o novo experimento melhore as estimativas do velho.

**Palavras-chave:** *Otimização, Análise sensorial, Testes triangulares .*

---

<sup>1</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [pcmribeiro@yahoo.com.br](mailto:pcmribeiro@yahoo.com.br)

<sup>2</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [jssbueno@dex.ufla.br](mailto:jssbueno@dex.ufla.br)

<sup>3</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [charlescfs@yahoo.com.br](mailto:charlescfs@yahoo.com.br)

# Análise da série de Produtividade de Soja em Municípios do Mato Grosso.

Gislaine Vieira Duarte <sup>14</sup>

Altemir da Silva Braga <sup>2</sup>

Vitor A. Ozaki <sup>3</sup>

**Resumo:** A modelagem das séries de produtividades agrícolas, na literatura, tem sido supostas por diferentes distribuições, tais como, Normal, Beta, Log-Normal, Skew-Normal. A maioria dos dados de produtividade agrícola apresentam assimetrias e/ou bimodalidade. Neste contexto, este trabalho teve por objetivo comparar as distribuições Normal, Skew-Normal e Rathie-Swamee na modelagem de séries de produtividade agrícola em municípios do estado do Mato Grosso no período de 1990 à 2012. Os resultados mostraram que o modelo que melhor se ajusta aos dados é o modelo Rathie-Swamee, capaz de capturar a multimodalidade dos dados.

**Palavras-chave:** *produtividade agrícola; Skew-normal; multimodalidade; Rathie-Swamee.*

## 1 Introdução

A modelagem da distribuição de produtividade tem sido um campo de pesquisa muito explorado nos últimos anos, este tema tem apresentado muitas discussões e controvérsias. Considerando a abordagem paramétrica, historicamente o modelo normal foi mais utilizado para modelar produtividade. No entanto, a distribuição normal é incapaz de captar assimetrias e bimodalidades que possam estar presentes nos dados.

Além da distribuição Normal outras distribuições paramétricas foram consideradas: distribuição **Gamma**, distribuição **Hiperbólica Inversa**, distribuição **Beta** e **Weibul**, distribuição **Normal-Assimétrica**.

No entanto, a distribuição paramétrica mais utilizada é a distribuição **Beta** [2],[6].

---

<sup>1</sup>LCE - ESALQ-USP: Escola Superior de Agricultura "Luis de Queiroz". Email: *gislaineveduarte@gmail.com*

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

<sup>2</sup>LCE - ESALQ-USP: Escola Superior de Agricultura "Luis de Queiroz". Email: *altemir@usp.br*

<sup>3</sup>LES - ESALQ-USP: Escola Superior de Agricultura "Luis de Queiroz". Email: *vaozaki@usp.br*

## 2 Objetivo

Este trabalho tem por objetivo modelar as séries de produtividade agrícola utilizando as distribuições paramétricas Normal, Skew-Normal [1] e Rathie-Swamee [9]. Esses novos modelos surgem como alternativa à distribuição Normal, pois apresentam forma flexível de densidade, podendo assim, capturar a assimetria ou bimodalidade dos dados.

## 3 Metodologia

As distribuições Normal e Skew-Normal são amplamente discutidas na literatura, portanto, a seguir faremos uma breve discussão da distribuição Rathie-Swamee (RS). No trabalho de [9] foi proposto uma generalização da distribuição logística com 3 parâmetros, cuja função de distribuição acumulada é dada por:

$$F(z) = \frac{1}{1 + e^{-z(a+b|z|^p)}}, a, b, p > 0$$

O modelo Rathie-Swamee  $RS(a, b, p)$  possui grande flexibilidade podendo assumir unimodalidade ou multimodalidade [4]. A função densidade de probabilidade é dada por:

$$f(x) = \frac{(a + b(1 + p)|z|^p)e^{-z(a+b|z|^p)}}{(e^{-z(a+b|z|^p)} + 1)^2}$$

Para possibilitar mais flexibilidade ao modelos RS, introduz-se os parâmetros de locação  $\mu$  e escala  $\sigma$ . Assim, se  $Z = \frac{Y-\mu}{\sigma}$  segue a distribuição  $RS(a, b, p)$  então  $Y$  segue a distribuição  $RS(a, b, p, \mu, \sigma)$  A Figura 1 apresenta a densidade da distribuição da Rathie Swamee(RS) para alguns valores dos parâmetros  $a, b, p \geq, \mu \in R, \sigma > 0$ . Pode-se observar na Figura 1 que quando o parâmetro  $a$  diminui a função densidade tende a zero no ponto  $t = 0$  causando o efeito de bimodalidade. Nota-se, também, o efeito do parametro  $b$  na contração e dilatação da base da função de densidade, causando uma diminuição da variabilidade da distribuição obtida. Nota-se o efeito do parâmetro  $p$  sobre o afastamento ou aproximação das possíveis modas.

## 4 Descrição dos dados

Os dados utilizados neste trabalho são de rendimento médio da lavoura temporária de soja (em grãos) e a unidade é quilogramas por hectare (Kg/ha) no período de 1990 à 2012 e foram coletados e disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) [7].

Antes de se ajustar qualquer modelo probabilístico para as séries de produtividade é ne-

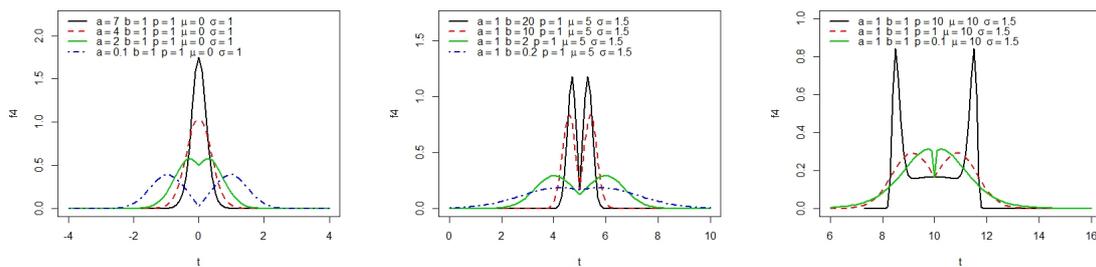


Figura 1: Gráfico da função densidade da Rathie-Swamee para alguns valores dos parâmetros.

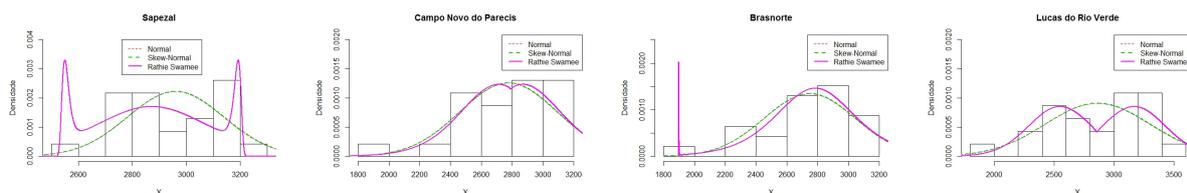


Figura 2: Distribuições Ajustadas para as séries estacionárias dos municípios de Mato Grosso.

cessário utilizar técnicas estatísticas que tornem os dados sem tendência, independentes e homocedásticos. A presença de tendência pode ser verificada através do teste de Phillips e Perron [8]. Para verificar se a série apresenta valores independentes, utiliza-se o gráfico da função de autocorrelação e o teste de Durbin Watson [5]. Após corrigido a tendência e a dependência das séries, estuda-se a hipótese de homocedasticidade através do teste de Breusch e Pagan [3]. Seleccionamos quatro municípios produtores de soja do estado do Mato Grosso que após a aplicação dos testes descritos acima apresentaram estacionariedade, independência e homocedasticidade. São eles: Sapezal, Campo Novo do Parecis, Lucas do Rio Verde e Brasnorte.

## 5 Resultados e discussões

Após a análise das estatísticas descritivas para os dados de produtividade de soja dos municípios analisados, observou-se que, todas as séries apresentaram mediana maior do que a média, e consequentemente, assimetria negativa. Ou seja, a curva da distribuição de probabilidade apresenta uma cauda mais longa à esquerda. Além disso, através do histograma, nota-se a presença de duas modas. A Figura 2 apresenta as densidades ajustadas pela distribuição Normal, Skewnormal e Rathie-Swamee para as séries de produtividade de soja escolhidas. Pode-se ver claramente que nas séries que apresentam bimodalidade evidentes, como a série de Sapezal e Lucas do Rio Verde, a distribuição RS consegue capturar esta bimodalidade. As estimativas dos modelos foram calculadas pelo método da máxima verossimilhança e o software utilizado foi o software R. A Tabela 1 apresenta as estimativas de máxima Verossimilhança para os parâmetros

dos modelos analisados e entre parênteses está o erro padrão das estimativas.

Tabela 1: Estimativas de máxima Verossimilhança para os parâmetros (erro padrão entre parênteses).

Municípios	Modelo	Estimativas				
Sapezal	Normal( $\mu, \sigma$ )	2960.3188 (37.3527)	179.1358 (26.4121)			
	Skew Normal( $\mu, \sigma, \lambda$ )	2959.8 (1910.12)	179.137(27.03)	0.0037(13.361)		
	RS( $a, b, p, \mu, \sigma$ )	2.57(3.59)	56.99(2064.28)	25.28(9.54)	2871.31(6.66)	376.63(519.51)
Lucas do Rio Verde	Normal( $\mu, \sigma$ )	2863.561 (91.22)	437.51 (64.50)			
	Skew Normal( $\mu, \sigma, \lambda$ )	2860.69 (2293.729)	437.524 (66.121)	0.0079 (6.565)		
	RS( $a, b, p, \mu, \sigma$ )	1.255 (13.90)	2.27(48.48)	0.933 (1.022)	2858.924 (76.7901)	754.64 (8317.38)
Brasnorte	Normal( $\mu, \sigma$ )	2748.56 (61.808)	296.423 (43.69)			
	Skew Normal( $\mu, \sigma, \lambda$ )	2747.27 (1634.698)	296.489 (44.21)	0.0051 (6.905)		
	RS( $a, b, p, \mu, \sigma$ )	5.147 (0.972)	0.01 (0.1405)	857.407(953.51)	2784.107 (61.88)	880.358 (63.4386)
Campo Novo de Parecis	Normal( $\mu, \sigma$ )	2776.35 (65.99)	316.50 (46.67)			
	Skew Normal( $\mu, \sigma, \lambda$ )	2775.03 (1682.17)	316.48 (47.14)	0.005 (6.656)		
	RS( $a, b, p, \mu, \sigma$ )	1.642(50.44)	0.328 (17.60)	0.745(1.22)	2790.73 (38.65)	356.43(10945.13)

Para a seleção do modelo que melhor represente os dados, utilizamos os critérios de informação Akaike modificado (AICc) e o critério de informação Bayesiano(BIC),no qual o menor valor é o preferido. Assim, a Tabela 2 nos apresenta o valor da estimativa de verossimilhança  $L(\hat{\theta})$  e os critérios de informação AICc e BIC. Portanto, o modelo que melhor se ajusta a todos os conjuntos de dados é o modelo Rathie-Swamee.

Tabela 2: Critérios de Informação

Municípios	Modelo	$L(\hat{\theta})$	AICc	BIC
Sapezal	Normal( $\mu, \sigma$ )	151.963	299.3259	297.6599
	Skew Normal( $\mu, \sigma, \lambda$ )	151.963	296.6628	294.5195
	RS( $a, b, p, \mu, \sigma$ )	143.35	283.1706	281.0225
Lucas do Rio Verde	Normal( $\mu, \sigma$ )	172.50	340.4038	338.7328
	Skew Normal( $\mu, \sigma, \lambda$ )	172.50	337.7406	335.5973
	RS( $a, b, p, \mu, \sigma$ )	171.2005	328.8716	326.7235
Brasnorte	Normal( $\mu, \sigma$ )	163.5514	322.502	320.83
	Skew Normal( $\mu, \sigma, \lambda$ )	163.5514	319.8396	317.6962
	RS( $a, b, p, \mu, \sigma$ )	159.32	305.114	302.96
Campo Novo do Parecis	Normal( $\mu, \sigma$ )	165.05	325.50	323.83
	Skew Normal( $\mu, \sigma, \lambda$ )	165.052	313.342	311.198
	RS( $a, b, p, \mu, \sigma$ )	164.7559	315.9824	313.8344

## 6 Conclusões

Nesta pesquisa foram avaliadas distribuições estatísticas alternativas para modelar dados de produtividade agrícola do IBGE. As séries temporais de produtividade de soja foram ajustadas considerando as distribuições Normal, Skew-Normal e Rathie-Swamee (RS) para 4 séries estacionárias de municípios do Estado do Mato Grosso. Conclui-se que a distribuição RS se ajusta melhor ao conjunto de dados.

Para próximas pesquisas, pode-se trabalhar com modelos de regressão e considerar a tendência natural das séries de produtividade.

## Referências Bibliográficas

- [1] AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian journal of statistics* (1985), 171–178.
- [2] BABCOCK, B. A., AND HENNESSY, D. A. Input demand under yield and revenue insurance. *American Journal of Agricultural Economics* 78, 2 (1996), 416–427.
- [3] BREUSCH, T. S., AND PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society* (1979), 1287–1294.
- [4] DE CASTRO GOMES, E. M. *Modelo Rathie-Swamee: aplicações e extensão para modelos de regressão*. PhD thesis, Escola Superior de Agricultura Luiz de Queiróz, 2013. Disponível em :<http://www.teses.usp.br/teses/disponiveis/11/11134/tde-26072013-105052/pt-br.php> Acessado em 20 de junho de 2014.
- [5] DURBIN, J., AND WATSON, G. S. Testing for serial correlation in least squares regression: I. *Biometrika* (1950), 409–428.
- [6] HENNESSY, D. A., BABCOCK, B. A., AND HAYES, D. J. Budgetary and producer welfare effects of revenue insurance. *American Journal of Agricultural Economics* 79, 3 (1997), 1024–1034.
- [7] IBGE. Instituto brasileiro de geografia e estatística. Acessado em 25 de junho de 2014 Disponível em: <http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=99&z=t&o=11>.
- [8] PHILLIPS, P. C., AND PERRON, P. Testing for a unit root in time series regression. *Biometrika* 75, 2 (1988), 335–346.
- [9] SWAMEE, P. K., AND RATHIE, P. N. Invertible alternatives to normal and lognormal distributions. *Journal of Hydrologic Engineering* 12, 2 (2007), 218–221.

# Relação dos Parâmetros do Método *Fuzzy k-Médias* com a Qualidade Estatística da Partição dos Dados

Nayara Luíza Sousa do Carmo<sup>1</sup>

Sueli Aparecida Mingoti<sup>2</sup>

Resumo: O uso de métodos de análise de *Cluster* para partição dos dados é muito comum, com aplicações nas mais diversas áreas. Existem vários métodos de agrupamento de dados, cada um com a sua particularidade e eficiência. Para utilizá-los o usuário precisa inicialmente, escolher a medida de dissimilaridade (ou distância) a ser usada para comparação dos elementos amostrais. Uma simples padronização pode alterar substancialmente a composição dos grupos da partição. Alguns métodos como o *Fuzzy k-Médias* por exemplo, exigem a pré-especificação de outros parâmetros além do número de grupos desejado para partição dos dados e medida de similaridade. Valores numéricos iguais de determinados parâmetros podem resultar em partições de qualidade estatística diferentes dependendo do conjunto de dados que está sendo analisado. O objetivo desse trabalho é mostrar através de estudo de casos, como a escolha dos valores numéricos dos parâmetros relacionados ao método *Fuzzy k-Médias* afeta a qualidade da partição final dos dados. Tem-se também como propósito avaliar qual é a abrangência da regra de escolha de parâmetros proposta em Schwammle e Jensen (2010). Em todos os casos, os agrupamentos obtidos pelo método *Fuzzy k-Médias* nas várias situações, são comparados com aqueles resultantes da aplicação do método não-hierárquico *k-Médias* e do hierárquico *Ward*, que são mais simples em termos de implementação. Os agrupamentos resultantes a partir da aplicação dos métodos de análise de *Cluster* citados e segundo variáveis medidas em cada elemento amostral, são comparados à um agrupamento prévio já existente nos dados, obtido por critérios externos às variáveis de agrupamento utilizadas na análise. As comparações são realizadas via medidas de concordância e medidas de qualidade estatística das partições obtidas.

**Palavras-chave:** *Análise de Cluster, Fuzzy k-Médias, k-Médias, Ward . Fuzzifier.*

*Projeto desenvolvido com apoio da PRPq/CNPq*

---

<sup>1</sup>DEST – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte – MG. Email: *nayluiza@est.grad.ufmg.br*

<sup>2</sup>DEST – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte – MG. Email: *sueliam@est.ufmg.br*

## **Classificação dos potenciais de viabilidade de lotes de sementes de mamona a partir análise visual da radiografias de mamonas**

**Rejane Corrêa da Rocha**<sup>1 4</sup>

**Thelma Sáfadi**<sup>2 4</sup>

**Maria Laene Moreira de Carvalho**<sup>3 4</sup>

O objetivo deste estudo foi implementar um método de classificação dos potenciais de viabilidade de lotes de sementes de mamona, a partir análise visual da radiografias de mamonas, utilizando técnicas de análise discriminante. Os dados foram obtidos de uma amostra de 400 sementes de mamona, da cultivar IAC-226, recém colhidas (safra 2009), sendo que as sementes foram identificadas individualmente e classificadas conforme seu potencial de viabilidade por três avaliadores diferentes. Foi ajustado o modelo linear de Fisher e calculada a função e o escore discriminante. Para avaliar a qualidade do ajuste foi utilizado o método da validação cruzada. O método proposto para a classificação da análise radiográfica mostrou-se eficiente, com uma proporção de acertos de 88%. Concluiu-se que o método proposto para a classificação pode ser uma alternativa viável para avaliar a qualidade de lotes de sementes de mamona.

**Palavras-chave:** *Análise de imagens, Análise discriminante, Teste de raios-X.*

---

<sup>1</sup>DEMAT- UFSJ: Universidade Federal de São João del-Rei, São João del-Rei -MG. Email: [rejane@ufsj.edu.br](mailto:rejane@ufsj.edu.br)

<sup>2</sup>DEX – UFLA: Universidade Federal de Lavras, Lavras-MG. Email [safadi@ufla.br](mailto:safadi@ufla.br)

<sup>3</sup>DAG – UFLA: Universidade Federal de Lavras, Lavras-MG. Email [mLaene@gmail.com](mailto:mLaene@gmail.com)

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

# Análise Exploratória Longitudinal: Um Estudo sobre o Impacto do Programa Bolsa Família no Atendimento Pré-Natal

Walmir dos Reis Miranda Filho<sup>1 4</sup>

Marcel de Toledo Vieira<sup>2 4</sup>

Ricardo da Silva Freguglia<sup>3 4</sup>

O Programa Bolsa Família (PBF) é um programa de transferência direta de renda criado em 2003 que beneficia famílias em situação de pobreza e de extrema pobreza. A Pesquisa de Avaliação do Impacto do Programa Bolsa Família (AIBF), um estudo em painel por amostragem complexa realizado nos anos de 2005 e 2009, é um importante instrumento para verificar se a referida política pública possui impacto nos mais diversos aspectos da vida de seus beneficiários. Sob o ponto de vista substantivo, nosso estudo tem como objetivo avaliar se o programa tem impacto positivo sobre o atendimento pré-natal em gestantes. Em termos metodológicos são investigados os impacto do plano amostral utilizado na pesquisa nas análises descritivas longitudinais conduzidas, no que diz respeito a estimativas pontuais e de precisão. O plano amostral empregado foi a amostragem em duas fases. Na primeira, foi adotada amostragem conglomerada para seleção de setores censitários, com estratificação geográfica e por tamanho do município. Na segunda, foi feita seleção de domicílios por amostragem estratificada simples. Além disso, a amostra da pesquisa foi dividida em três grupos, sendo um grupo de tratamento, formado por domicílios beneficiários diretos do PBF, e dois de controle, o primeiro representado por domicílios beneficiários de outros programas do governo federal que não o Bolsa Família ou presente no Cadastro Único e o segundo, por domicílios sem qualquer benefício e não cadastrados.

**Palavras-chave:** *Bolsa Família, Pré-natal, Efeitos do Plano Amostral, AIBF.*

---

<sup>1</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora - MG. Email: [walmir.reis.miranda.filho@gmail.com](mailto:walmir.reis.miranda.filho@gmail.com)

<sup>2</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora - MG. Email: [marcel.vieira@ice.ufjf.br](mailto:marcel.vieira@ice.ufjf.br)

<sup>3</sup> Fac. de Economia – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora - MG. Email: [ricardo.freguglia@ufjf.edu.br](mailto:ricardo.freguglia@ufjf.edu.br)

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

# Gráficos de Controle Multivariados para a Melhoria da Qualidade do Processo Produtivo de uma Panificadora

Raphael Henrique Teixeira da Silva<sup>1 3</sup>

Daniela Carine Ramires de Oliveira<sup>2 3</sup>

**Resumo:** As empresas de diversos segmentos existentes no mercado têm tendência em controlar seus processos produtivos, a fim de reduzir o desperdício, retrabalho e a variabilidade dos produtos. Neste sentido, este trabalho teve como objetivo aplicar métodos estatísticos para a melhoria da qualidade do pão de sal produzido por uma padaria situada no interior do estado de Minas Gerais. Mais especificamente, foram utilizados os gráficos de controle de qualidade multivariados, por haver três características essenciais para a produção desse produto. Além disso, os gráficos foram gerados em um software estatístico livre, não gerando custos adicionais para a empresa. O trabalho proposto teve grande impacto na empresa, pois possibilitou identificar os problemas e propor melhorias com relação ao *layout* do setor, da matéria-prima e do produto.

**Palavras-chave:** *Gráfico|S|, Gráfico  $T^2$  de Hotelling, Decomposição da estatística  $T^2$ .*

## 1 Introdução

A alta competitividade existente no mercado é um fator cada vez mais presente na realidade das empresas. Para tanto, pode-se dizer que o controle estatístico da qualidade é uma forma eficaz para se controlar um processo garantindo a melhoria contínua da produção. Os gráficos de controle se destacam como uma das principais ferramentas estatísticas que garantem a estabilidade de um processo. O principal objetivo desses gráficos é detectar causas especiais (inaceitáveis) existentes em um processo. Os gráficos de controle univariados são os mais comuns na realidade das empresas e servem de grande auxílio para o monitoramento *on-line* da produção. Entretanto, quando se avalia diversas características simultaneamente e, as mesmas estão correlacionadas, esses gráficos, em geral, não são eficientes. Para essas situações, os gráficos de controle multivariados (GCM) são mais

---

<sup>1</sup> UFSJ: Universidade Federal de São João del Rei, São João del Rei-MG. Email: [raphaelhts@hotmail.com](mailto:raphaelhts@hotmail.com)

<sup>2</sup> DEMAT – UFSJ: Universidade Federal de São João del Rei, São João del Rei-MG. Email: [ramires.daniela@gmail.com](mailto:ramires.daniela@gmail.com)

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

eficazes. Dentre os existentes, pode ser utilizado o gráfico da variância generalizada ( $|S|$ ), para monitorar a variabilidade do processo e o gráfico  $T^2$  de Hotelling, para monitorar a média.

Lima *et al.* (2012) utilizaram técnicas estatísticas multivariadas para aperfeiçoar o processo de monitorar e controlar a qualidade numa empresa que produz garrafas plásticas, sendo uma delas o gráfico  $T^2$  de Hotelling aplicado a um conjunto de dez características de qualidade. Oyeyemi (2011) apresenta o gráfico  $T^2$  de Hotelling para o controle do processo de produção de um detergente, considerando quatro características. Rojas *et al.* (2011) utilizaram o  $T^2$  de Hotelling para avaliar a diferenciação morfométrica entre duas espécies que são parasitas de um rebanho de cabras.

O objetivo deste trabalho foi apresentar um estudo de caso em uma panificadora, aplicando os gráficos de controle de qualidade citados anteriormente para a melhoria do processo de produção da massa do pão de sal. Para geração dos gráficos foi utilizado o software livre R.

## 2 Metodologia

A empresa estudada é uma padaria situada em Ponte Nova-MG. A mesma produz vários produtos para comercialização no mercado, entretanto, o produto mais vendido é o pão de sal. Portanto, a produção da massa do pão de sal é realizada com mais frequência, sendo cinco vezes em cada dia (três vezes de manhã e duas vezes à tarde). Com isso, a empresa resolveu aplicar métodos de controle para acompanhar esse processo, para reduzir variabilidade, desperdício e retrabalho.

Devido aos objetivos citados anteriormente foi aplicado uma folha de verificação para a coleta de dados. A mesma consistiu de informações da produção da massa (concentração de cada um dos ingredientes). As amostras para este estudo foram coletadas no mês de Novembro de 2013. As características de estudo foram: farinha, açúcar e reforçador. A escolha dessas características se deve ao custo desses produtos e também por serem os ingredientes com mais alta porcentagem de adição na massa. A receita padrão da padaria de estudo consiste em utilizar 1500g de farinha, 180g de açúcar e 150g de reforçador.

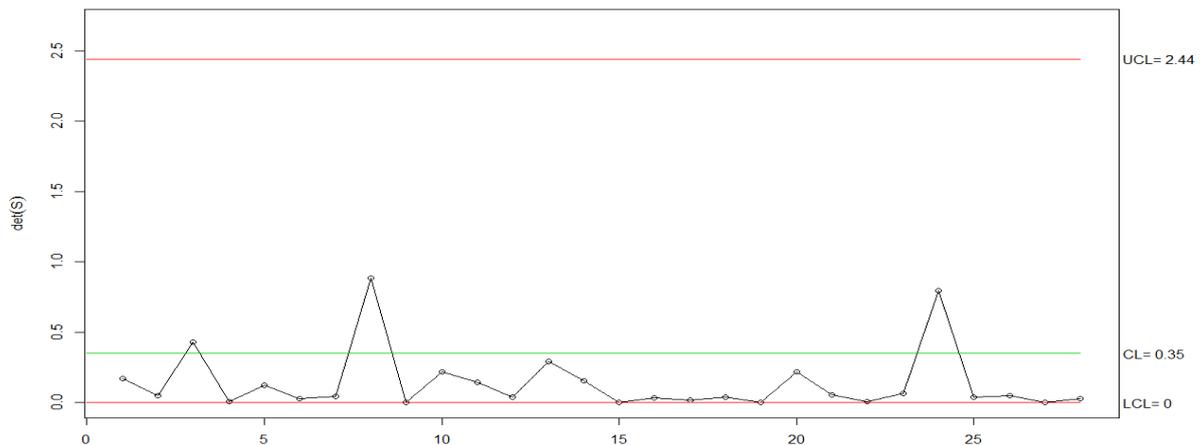
Inicialmente foi feita a verificação da normalidade multivariada e o teste da correlação entre as características. Como ambos foram confirmados, foram gerados os gráficos multivariados: variância generalizada (gráfico  $|S|$ ) e  $T^2$  de Hotelling. Por último, a decomposição da estatística  $T^2$  de Hotelling, para o estudo das causas especiais do processo.

O embasamento teórico para a construção dos gráficos de controle e o método da decomposição da estatística  $T^2$  de Hotelling foi obtido de Montgomery (2009).

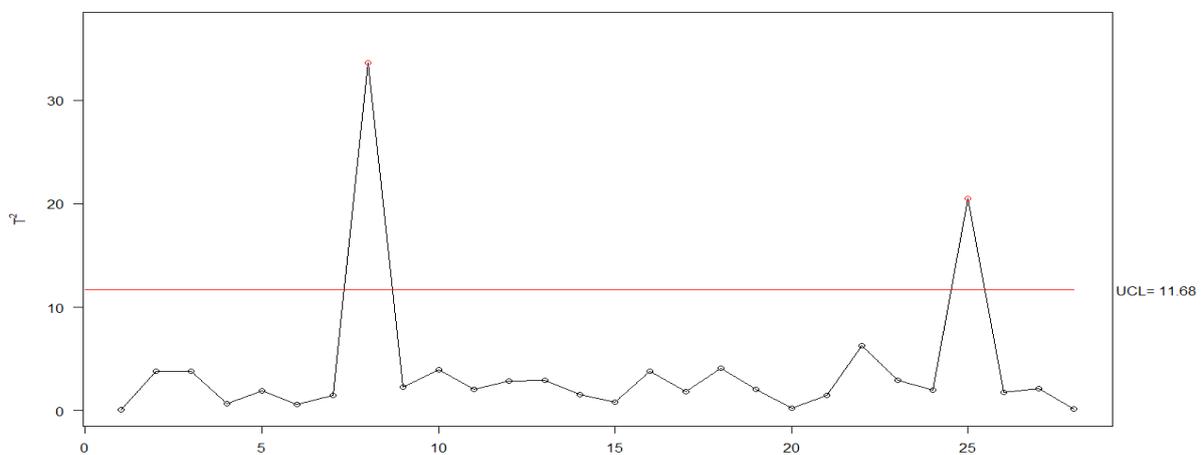
Os gráficos foram elaborados com o auxílio do software livre R: pacote MSQC (*Multivariate Statistical Quality Control*), para gerar os multivariados e realizar a decomposição. Para a verificação da suposição de normalidade multivariada foi utilizado o pacote *mvnormtest* (*Normality test for multivariate variables*), que faz o teste de *Shapiro Wilk*. Além disso, foi adotado um nível de significância de 0,05.

### 3 Resultados e Discussões

Com o uso das folhas de verificação e do acompanhamento do processo de produção do pão de sal, foram identificadas variações e a necessidade de melhoria. Portanto, foi realizado o teste de correlação entre as características da qualidade (farinha, açúcar e reforçador) e a suposição da normalidade multivariada, indicando que as variáveis tinham correlação significativa e que possuíam distribuição normal multivariada, considerando um nível de significância de 5%. Por isso, foram gerados os gráficos de controle multivariados (gráfico  $|S|$  e  $T^2$  de Hotelling), para avaliar as características simultaneamente (ver Figuras 1 e 2).



**Figura 1.** Gráfico da variância generalizada para concentrações de Farinha, Açúcar e Reforçador.



**Figura 2.** Gráfico  $T^2$  de Hotelling para as concentrações de Farinha, Açúcar e Reforçador.

Observando a Figura 1, com relação à variabilidade, o processo está sob controle estatístico. Contudo, observando a Figura 2 o GCM  $T^2$  de Hotelling apresentou dois pontos acima do limite superior de controle. As amostras 08 e 25 correspondentes aos dias 10 e 27 de Novembro, respectivamente, indicaram que nesses dias ocorreram algumas causas especiais que impactaram o processo de forma negativa. Portanto, o próximo passo foi realizar a decomposição da estatística  $T^2$  para descobrir qual característica da qualidade influenciou o processo de forma inadequada.

O software R com os recursos da biblioteca MSQC apresenta os valores da decomposição da estatística  $T^2$  para as amostras que saíram do limite superior de controle.

**Tabela 1.** Decomposição da estatística  $T^2$  de Hotelling para as amostras 08 e 25.

<b>Amostra 08</b>	<b><math>d_i</math></b>	<b>Amostra 25</b>	<b><math>d_i</math></b>
Farinha	27,0243	Farinha	12,2637
Açúcar	4,9798	Açúcar	6,5748
Reforçador	18,1934	Reforçador	7,5166

Fonte: Resultados obtidos da decomposição das amostras 08 e 25.

Observando os valores de  $d_i$  das amostras 08 e 25 (ver Tabela 1), os maiores valores foram para a concentração de Farinha. Uma investigação foi realizada na empresa, e pôde ser constatado que no dia 10/11, o recipiente de plástico utilizado para pegar a farinha quebrou, e isso prejudicou o serviço do funcionário que utilizou outro recipiente que não era adequado para este processo, isto é, era maior do que o utilizado anteriormente. Já no dia 27/11 a entrega da farinha atrasou. Segundo os proprietários da empresa, a entrega era para ser feita no dia 26/11 às 17h. Contudo, a farinha só chegou no dia 27/11 na parte da manhã, por volta das 10h30min. Isso acarretou em um atraso no processo, porque os funcionários precisaram produzir o pão e receber o estoque de farinha, simultaneamente.

## 4 Conclusões

O trabalho proposto teve grande impacto na empresa estudada, afinal possibilitou a empresa uma técnica mais robusta para o controle de um de seus principais processos. Aplicando os gráficos de controle multivariados, foi possível detectar dois pontos fora do limite superior de controle e consequentemente, descobrir e corrigir os problemas encontrados.

Foi proposto que houvesse uma inspeção dos equipamentos utilizados na produção pelo menos uma vez por semana, como também comprar utensílios com maior resistência para garantir uma maior durabilidade. Com relação ao atraso da farinha, os proprietários manterão um estoque reserva maior de farinha, para que o suprimento base deles não falte.

É importante ressaltar que, o trabalho contou com o auxílio de um software livre, que não gerou para a empresa custos fora do orçamento.

## Referências

- [1] LIMA, M. B. F.; SANTOS, R. L. S.; SOUZA, E. L.; SILVA, L. B. Aplicações do Controle Estatístico Multivariado da Qualidade: Controlando Garrafeiras Plásticas numa Empresa Industrial. *Cadernos do IME – Série Estatística*. ISSN impresso 1413-9022 / ISSN on-line 2317-4535 - v. 32, p. 01 - 14, 2012.
- [2] MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. 6 ed. United State of America: John Wiley & Sons, 2009.
- [3] OYEYEMI, G. M. Principal Component Chart For Multivariate Statistical Process Control. *The Online Journal of Science and Technology, Turkey*, v.1, n. 2, p.22 – 31, April 2011.
- [4] R Development Core Team. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2013. Disponível em:<<http://www.r-project.org>>. Acesso em: 15 Set. 2013.
- [5] ROJAS N.; La O M.; ARECE J.; CARRION M.; PEREZ K.; MARTIN C. S.; VALERIANO P.; RAMIREZ W. Identificación y caracterización de especies de *Haemonchus* en caprinos del valle del Cauto en Granma. *Revista Electrónica de Veterinaria*, v. 13, n. 1, 2011.

# A distribuição Dagum transmutada

Jailson de Araujo Rodrigues<sup>1</sup>

Ana Paula Coelho Madeira Silva<sup>2</sup>

As distribuições de probabilidade tem sido extensivamente utilizadas na modelagem de dados em diferentes domínios científicos. Esse fato faz com que sua teoria esteja sempre em pauta e novas distribuições sejam desenvolvidas.

O modelo Dagum foi originalmente proposto para estudar dados de renda. Ao longo dos anos, diversos autores passaram a estudar e aplicar esse modelo em diferentes contextos. Isso levou ao surgimento de generalizações e formas modificadas desse modelo.

Neste trabalho, é introduzida uma distribuição com quatro parâmetros denominada Dagum transmutada. Algumas das principais propriedades estruturais dessa distribuição são apresentadas. Em particular, são encontradas expressões para os momentos, os desvios médios e as curvas de Bonferroni e Lorenz. A estimação dos parâmetros é discutida utilizando os métodos dos momentos e da máxima verossimilhança. O ajuste da distribuição a um conjunto de dados de sobrevivência demonstrou a versatilidade do modelo.

**Palavras-chave:** *Dagum, Momentos, Estimadores de Máxima Verossimilhança.*

---

<sup>1</sup>IFBA: Instituto Federal da Bahia, Feira de Santana-BA. Email: [jailsondearaujo@yahoo.com.br](mailto:jailsondearaujo@yahoo.com.br)

<sup>2</sup>UFSJ: Universidade Federal de São João Del Rei, Sete Lagoas. Email: [apcmadeira@hotmail.com](mailto:apcmadeira@hotmail.com)

# Medição e Análise no CMMI com Metodologia Seis Sigma e ISO/IEC 15939

**Charles Shalimar Felipe da Silva<sup>1</sup>**

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

**Paulo César Moraes Ribeiro<sup>3</sup>**

O CMMI - Capability Maturity Model Integration, ou Integração de Modelos de Maturidade de Capacidade é um dos modelos de melhoria de processos de software mais bem sucedidos. Este trabalho se propõe a conjugar a metodologia Seis Sigma com a norma ISO/IEC 15939 para atender as especificações demandadas pelo CMMI para o sistema de medição e galgar os degraus de melhoria indicados pelos níveis daquele modelo com o uso de métodos estatísticos conforme a metodologia Seis Sigma.

A metodologia Seis Sigma busca definir os problemas a melhorar, medir e analisar informação, e controlar processos ou produtos com o fim de alcançar etapas ótimas.

A norma ISO/IEC 15939 consiste num processo de mensuração e análise de instâncias de projeto, riscos e custos, para projetos de desenvolvimento de software. Seu modelo de informação fornece uma estrutura que relaciona necessidades de informação e conjunto de medidas, enquanto o processo de medição estabelece, sustenta, planeja, realiza e avalia a medição.

Num estudo de caso em projetos de uma empresa de software verificou-se a necessidade de medição de atributos, inicialmente desconsiderados pela organização, mas que, pela análise proposta, mostram-se relevantes.

Os argumentos desenvolvidos neste trabalho demonstram a contribuição das metodologias estatísticas para tomada de decisões “certeiras” para melhoria de processos de software.

**Palavras-chave:** *CMMI, Seis Sigma, Medição.*

---

<sup>1</sup> DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [charlesfsilva@yahoo.com.br](mailto:charlesfsilva@yahoo.com.br)

<sup>2</sup> DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [marcelo.oliveira@dex.ufla.br](mailto:marcelo.oliveira@dex.ufla.br)

<sup>3</sup> DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [pcmribeiro@yahoo.com.br](mailto:pcmribeiro@yahoo.com.br)

# Identificação de um processo de Markov em experimentos quânticos com spin de elétrons e em polarização de fótons

Felipe Andrade Velozo<sup>1 3</sup>

José Alberto Casto Nogales Vera<sup>2 3</sup>

Marcelo Silva de Oliveira<sup>2 3</sup>

Lucas Monteiro Chaves<sup>2 3</sup>

Devanil Jaques de Souza<sup>2 3</sup>

Em experimentos quânticos em que um elétron passa por uma sequência de aparatos de Stern-Gerlach, bem como no experimento em que um fóton passa por uma sequência de polarizadores, observa-se que esses experimentos podem ser modelados por um processo de Markov. Tais experimentos aparecem em artigos que tratam sobre as dificuldades de se estabelecer relações entre o Cálculo de Probabilidade e a Teoria Quântica (artigos sobre “variáveis ocultas”, ou melhor, variáveis aleatórias em que não se tem controle sobre os valores assumidos por elas). Neste trabalho tem-se o objetivo apenas de apresentar a modelagem do problema através de um processo de Markov, uma vez que não se encontra na literatura de Teoria Quântica tal explicitação, apesar do experimento ter suas probabilidades bem definidas. Para isso, o trabalho se apoia em resultados apresentados na literatura de Teoria Quântica, para examiná-los de forma a orientar a identificação do processo envolvido. Com isso, conclui-se que o experimento em questão possui uma modelagem condizente no Cálculo de Probabilidade, possibilitando um questionamento acerca do uso desses experimentos como base para refutação à possibilidade de se empregar o Cálculo de Probabilidade em todos os sistemas quânticos. Deste trabalho levanta-se a seguinte questão: um sistema quântico que pôde ser modelado pelo Cálculo de Probabilidade pode ser usado para refutar o próprio Cálculo de Probabilidade, tornando necessário um Cálculo de Probabilidade Quântico?

**Palavras-chave:** *Processo de Markov, Spin de elétrons ao passar por diversos aparatos de Stern-Gerlach, Polarização de fótons ao passar por diversos polarizadores.*

---

<sup>1</sup>Doutorando do DEX – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [felipe.andrade.velozo@mail.ru](mailto:felipe.andrade.velozo@mail.ru)

<sup>2</sup>Professor do DEX – UFLA: Universidade Federal de Lavras, Lavras-MG

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

# Uma nova generalização da distribuição Lomax com aplicação em Análise de Sobrevivência

Jailson de Araujo Rodrigues <sup>1</sup>

Ana Paula Coelho Madeira Silva<sup>2</sup>

Ângela Lima da Silva<sup>3</sup>

A distribuição Lomax ou Pareto tipo II como também é conhecida foi originalmente proposta para analisar dados de insucesso empresarial. Na última década foram apresentadas diversas generalizações dessa distribuição. Neste trabalho, é introduzida uma generalização do modelo Lomax denominada Kumaraswamy exponenciada-Lomax. Além da distribuição Lomax, o modelo inclui as distribuições Lomax exponenciada, Lomax generalizada exponenciada e Kumaraswamy Lomax. Algumas das principais propriedades estruturais dessa distribuição são apresentadas. A estimação dos parâmetros do modelo é discutida utilizando os métodos dos momentos e da máxima verossimilhança. O potencial do modelo proposto é demonstrado no ajuste de um conjunto de dados de sobrevivência.

**Palavras-chave:** *Critério de Informação de Akaike, Distribuição Lomax, Distribuição Kumaraswamy, Momentos, Método da máxima verossimilhança.*

---

<sup>1</sup>IFBA: Instituto Federal da Bahia, Feira de Santana-BA. Email: [jailsondearaujo@yahoo.com.br](mailto:jailsondearaujo@yahoo.com.br)

<sup>2</sup>UFSJ: Universidade Federal de São João Del Rei, Sete Lagoas. Email: [apcmadeira@hotmail.com](mailto:apcmadeira@hotmail.com)

<sup>3</sup>UFBA: Universidade Federal da Bahia, Salvador.

# ESTUDOS DE DESCRITORES MORFOLÓGICOS DE MACAÚBA NA FASE JUVENIL

**Ronaldo Machado Junior<sup>1,2</sup>**

**Sylas Clemente de Oliveira<sup>1</sup>**

**Guilherme Ferreira Alves<sup>1</sup>**

**Luiz Henrique de Oliveira Dianin<sup>1</sup>**

**Éder Matsuo<sup>1,2</sup>**

Frutos de macaúba (*Acrocomia aculeata*) têm alto potencial para produção de óleos. No entanto é importante o desenvolvimento e aprimoramento de programas de melhoramento genético. Objetivou-se avaliar progênies de macaúba por meio de descritores morfológicos visando direcionar futuros trabalhos. O trabalho foi desenvolvido, em campo na Universidade Federal de Viçosa – *Campus* Rio Paranaíba. Foram avaliadas 36 progênies de macaúba, distribuídas no delineamento em blocos casualizados com três repetições, sendo duas plantas por parcela. Doze meses após o transplântio, as plantas foram avaliadas quanto à altura (AP), diâmetro do estipe (DE), projeção da copa (PC) e número de folhas por ano (NF). Os dados foram submetidos ao teste F na análise de variância, o qual apresentou efeito significativo para os descritores. Estimou-se o intervalo de confiança com 95% de probabilidade para médias dos descritores e realizou análise de repetibilidade. Os intervalos de confiança obtidos foram  $39,3 \pm 1,8$  cm (AP),  $11,9 \pm 0,6$  mm (DE),  $40,6 \pm 2,0$  cm (PC) e  $2,0 \pm 0,1$  (NF). Pela análise de repetibilidade com duas medições para DE e NF e três para AP e PC, obteve-se 95% de confiabilidade. Assim, puderam-se obter estimativas importantes para próximos estudos.

**Palavra chave:** *Acrocomia aculeata*, Melhoramento, Biometria.

---

1 Universidade Federal de Viçosa – *Campus* Rio Paranaíba. Emails: [ronaldo.machado@ufv.br](mailto:ronaldo.machado@ufv.br), [sylas.oliveira@ufv.br](mailto:sylas.oliveira@ufv.br), [guilherme.f.alves@ufv.br](mailto:guilherme.f.alves@ufv.br), [luiz.dianin@ufv.br](mailto:luiz.dianin@ufv.br) e [edermatsuo@ufv.br](mailto:edermatsuo@ufv.br)

2 Agradecimento à FAPEMIG e PÉTROBRAS pelo apoio financeiro.

# Perfil dos formandos em Engenharia Civil da Universidade Federal de Uberlândia

Jéssica Viviana Martins Mendes <sup>1,4</sup>

André Luiz Rodrigues Souza <sup>2,4</sup>

Janser Moura Pereira <sup>3,4</sup>

Atualmente as formas de ingressar na UFU são: SISU (Sistema de Seleção Unificada); PAAES (processo seletivo de composto de três etapas); Vestibular; Portador de diploma; Transferência Interna/Externa; Convênio e Mobilidade Internacional/Nacional. Tomou-se como base o banco de dados de 2012 a 2013 cedido pela Diretoria de Administração e Controle Acadêmico da Universidade Federal de Uberlândia (DIRAC/UFU). No presente trabalho foi aplicado um teste de comparações múltiplas entre proporções para verificar se há diferença significativa nas proporções das formas de ingresso. Possibilitando assim traçar um perfil do discente desse curso. Concluiu-se que existe diferença significativa entre as proporções das formas de ingresso, sendo a maior proporção dos ingressantes por meio do vestibular (77,23%), seguido pelo Paies (18,70%), transferência externa (2,44%) e transferência interna (1,63%). As novas formas de ingresso, PAAES e SISU, não foram avaliadas pois não existiam quando os formandos de 2012 e 2013 ingressaram na Universidade Federal de Uberlândia.

**Palavras- chave:** *Comparações múltiplas assintóticas, Proporções binomiais, Ingressantes.*

---

<sup>1</sup> FAMAT –UFU: Universidade Federal de Uberlândia ,Uberlândia -MG.  
Email : [jessicaviviana@est.ufu.br](mailto:jessicaviviana@est.ufu.br)

<sup>2</sup> FAMAT –UFU: Universidade Federal de Uberlândia ,Uberlândia -MG.  
Email : [arodrigues.est.ufu@gmail.com](mailto:arodrigues.est.ufu@gmail.com)

<sup>3</sup> FAMAT –UFU: Universidade Federal de Uberlândia ,Uberlândia -MG.  
Email : [jansermp@gmail.com](mailto:jansermp@gmail.com)

<sup>4</sup> Agradecimento à FAPEMIG pelo apoio financeiro e à DIRAC pela concessão dos dados.

# Análise do Aprendizado por Reforço na Simulação de Futebol de Robôs via Regressão Logística

André Luiz Carvalho Ottoni<sup>1 3</sup>

Marcos Santos de Oliveira<sup>2 3</sup>

O aprendizado por reforço é um formalismo da Inteligência Artificial que permite a um agente aprender a partir da sua interação com o ambiente no qual ele está inserido. Quanto mais complexo o ambiente, isto é, quanto maior o número de ações ou a quantidade de agentes, maior é a capacidade computacional necessária para resolver o problema. Dessa forma, verificar o processo de convergência de um sistema de aprendizado por reforço é importante a fim de evitar esforços computacionais desnecessários. A categoria de simulação 2D da *RoboCup* (competição de robótica internacional) simula partidas de futebol de robôs autônomos. O objetivo desse trabalho foi verificar a análise de convergência do aprendizado por reforço via modelos de regressão logística. Procurou-se avaliar a convergência do algoritmo *Q-learning* aplicado em um time de futebol de robôs simulado. Na etapa experimental foram simuladas 600 partidas de futebol de robôs na plataforma de *RcSoccerSim* da *RoboCup*. Ao final de cada partida foi armazenado o resultado final do jogo (vitória ou não vitória) e número de toques na bola de cada robô (jogador) na partida. Com as análises dos modelos de regressão logísticos construídos foi possível verificar o ponto de estabilização do aprendizado por reforço considerando a influência de cada setor (defesa, meio-de-campo e ataque) no resultado no final da partida.

**Palavras-chave:** *Robótica, Inteligência Artificial, Convergência.*

---

<sup>1</sup>DEPEL – UFSJ: Universidade Federal de São João del-Rei, São João del-Rei-MG. Email: [andreettoni@ymail.com](mailto:andreettoni@ymail.com)

<sup>2</sup>DEMAT – UFSJ: Universidade Federal de São João del-Rei, São João del-Rei-MG. Email: [mso@ufsj.edu.br](mailto:mso@ufsj.edu.br)

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

# Comparativo do Desempenho de Análise de RNA-Seq em Experimentos com Diferentes números de Fatores e Repetições

Vladimir Barbosa Carlos de Souza<sup>1 3</sup>

Luiz Alexandre Peternelli<sup>2 3</sup>

A tecnologia RNA-Seq, baseada em tecnologias de Sequenciamento de Nova Geração (NGS), mostrou-se ser revolucionária para o estudo de expressão gênica. Porém, mais estudos na literatura sobre a análise de dados de RNA-Seq são necessários, até mesmo porque trata-se de um método de elevado custo. Devido a este alto custo, é importante o aproveitamento das amostras disponíveis para concluir sobre mais fatores. Este trabalho tem como objetivo realizar um comparativo do desempenho da análise de identificação de genes diferencialmente expressos (DEGs) em experimentos fatoriais com diferentes números de fatores e repetições. Para a análise, foram simulados vários conjuntos de dados de RNA-Seq, provenientes de experimentos fatoriais com diferentes números de fatores e repetições, mas que, em todas as simulações, o número de amostras é o mesmo, ou seja, foram simulados experimentos com o mesmo custo. Para realizar a simulação, foi utilizado o pacote TCC, desenvolvido para o software livre R, assim como a normalização dos dados e identificação de DEGs também foram realizadas com o uso deste pacote. E por último, a comparação do desempenho da análise feita em cada experimento foi realizada utilizando-se curvas ROC (Receiver Operating Characteristics). Percebeu-se que, ao aumentar o número de fatores e diminuir o número de repetições, o desempenho da análise vai diminuindo até certo ponto e, depois dele, o desempenho aumenta.

**Palavras-chave:** *Experimentos fatoriais, expressão gênica, curvas ROC*

---

<sup>1</sup>Mestrando em Estatística Aplicada e Biometria, Universidade Federal de Viçosa. Email: vladimir.souza@ufv.br

<sup>2</sup>Professor Associado, Departamento de Estatística, Universidade Federal de Viçosa. Email: peternelli@ufv.br

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

# Aplicação do método de Bland-Altman para análise de concordância de estimativas para filtração glomerular no diagnóstico de Doença Renal Crônica

Guilherme William Marcelino<sup>1 3</sup>

Fernando A.B. Colugnati<sup>2 3</sup>

A taxa de filtração glomerular (TFG) é usada para diagnosticar doenças renais crônicas (DRC) a TFG é estimada a partir da Creatinina Sérica. Existem duas fórmulas para estimar a TFG a partir da creatinina medida no sangue, a MDRD e a CKDEPI. A MDRD foi originalmente desenvolvida com base nos dados do estudo Modification of Diet in Renal Disease (MDRD) em pacientes com DRC e não incluiu indivíduos saudáveis, enquanto a CKDEPI que foi desenvolvida pelo o grupo Chronic Kidney Disease Epidemiology Collaboration recentemente que inclui indivíduos com e sem doença DRC. Para comparar ambas, usamos o método de Bland-Altman em uma amostra de 200 usuários do CDHM-JF com uma amostra aleatória selecionada a partir do prontuário eletrônico. Observou-se uma correlação de 0.98 entre as estimativas e um viés estimado de 3.1(IC 2.6-3.7) do CKDEPI em relação ao MDRD. Pelo gráfico de diferenças x médias, observamos que CKDEPI superestima em relação ao MDRD e a variância das diferenças não é constante, quanto maior é a TFG maior é a variância das diferenças. Esta comparação é importante pois, se um mesmo paciente pode ser diagnosticado de formas diferentes, é necessário que se avalie estatisticamente estas diferenças.

**Palavras-chave:** *MDRD, CKDEPI, Bland-Altman, DRC, TFG, diagnostico.*

---

<sup>1</sup>Dep.de Estatística/ICE –UFJF. e-mail: guilherme1marcelino@gmail.com

<sup>2</sup>Dep.de Clínica Médica/FaMed -UFJF. e-mail: fcolugnati@gmail.com

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

# Espaços de Núcleos de Reprodução e o Modelo SIR

Estela Costa Ferreira<sup>1 3</sup>

José Claudinei Ferreira<sup>2</sup>

3

O estudo de doenças infecciosas é imprescindível para o controle de epidemias. Com a finalidade de descobrir as causas e caracterizar cada epidemia, pesquisadores aliaram a modelagem matemática ao controle biológico. Diversos problemas de análise matemática e de física nos conduzem a equações integrais, em particular para equações do tipo Volterra. O modelo SIR, proposto por Kermack e McKendrick em 1927, se destaca entre os modelos de epidemiologia e pode ser escrito como uma equação integral de Volterra não linear. Este modelo leva em consideração três subgrupos da população: S, das pessoas suscetíveis a doença; I, das pessoas infectadas e R, das pessoas removidas do sistema. A resolução de uma equação integral não linear, por meio de métodos diretos ou analíticos, não é algo simples, sendo que muitas vezes conseguimos apenas demonstrar teoremas sobre condições de existência de tais soluções. Logo, um método muito comum de lidar com tais equações é fazer aproximações para as soluções, ou usar métodos de linearização e estudar o seu comportamento. Faremos um estudo da teoria de espaços de núcleo de reprodução, uma vez que as propriedades e a estrutura de espaço de Hilbert podem ser usadas para encontrar uma aproximação para a solução de certas equações lineares. Esses espaços são amplamente utilizados para a modelagem de fenômenos com uma grande quantidade de dados, como em modelos que tratam do processo de produção de proteínas, técnicas de inteligência artificial, etc.

**Palavras-chave:** *Equações integrais, espaços de núcleos de reprodução, modelagem matemática, modelo SIR..*

---

<sup>1</sup>Mestrando em Estatística Aplicada e Biometria. Universidade Federal de Alfenas, alfenas-MG. Email: *estela\_cf@hotmail.com*

<sup>2</sup>Orientador, Faculdade de Matemática. Universidade Federal de Uberlândia, Uberlândia-MG. Email: *claudinei@famat.ufu.br*

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

# Comparação de modelos para ajuste de curvas de lactação de diferentes raças Bovinas

Rafaela Maia Moreira<sup>1</sup>

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

**Resumo:** A produção leiteira de um animal ou rebanho é normalmente representada graficamente por curvas ajustadas segundo um modelo que possibilite encontrar parâmetros que possam embasar critérios de seleção, bem como estimar seu desempenho. Foram estudados quatro modelos estatísticos: Linear, Quadrático, Linear Hiperbólico e de Wood, sendo o último mais utilizado. Para o ajuste utilizou-se rotina no software R, do qual foram obtidos os parâmetros estimados, o valor-p referente a cada ajuste e os respectivos gráficos. Os dados utilizados foram de produções leiteiras de 43 vacas de primeira lactação sendo 17 da raça Gir, 20 da Holandesa e 6 da Jersey, realizando-se o ajuste por animal e por raça. Em todos os casos, avaliando-se o coeficiente de determinação, destacaram-se os modelos de Wood e linear Hiperbólico, tendo ainda melhor desempenho para a raça Holandesa. Para os dois modelos foram calculados o pico de produção de leite, através da derivada da função, a produção total de leite pela integral de 0 até 305 dias após o parto e a persistência na lactação, dada pela relação entre a produção total e o valor máximo da produção leiteira. Notou-se também que os melhores ajustes ocorreram quando havia maior disponibilidade de dados, facilitando assim a caracterização dos animais e dos rebanhos.

**Palavras-chave:** *Modelo de Wood, Função Linear hiperbólica, pico de lactação, persistência.*

---

<sup>1</sup> DEZOO- UFSJ. e-mail: rafaelammoreira13@gmail.com

<sup>2</sup> DEZOO- UFSJ. e-mail: carlabrighenti@ufsj.edu.br

# Esperança de vida: Análise de dados do IBGE e DATASUS da população de Ouro Preto no ano de 2010.

Numiá Freitas Gomes<sup>1</sup>

Vanessa Miranda Oliveira<sup>2</sup>

Graziela Dutra Rocha Gouvêa<sup>3</sup>

A esperança de vida representa o número médio de anos vividos por uma população e apresenta grande importância para conhecimento e estudo das condições sociais de uma população. A Tábua de Mortalidade (Tábua de Vida) mecanismo, é uma tabela utilizada para calcular as probabilidades de vida e morte de uma população, em função da idade (Rollet, 2007). São criadas a partir de dados provenientes de Censos Populacionais, entidades oficiais do país a que se referem e registro civil e reflete a população num determinado período de tempo. O objetivo deste trabalho é estimar a esperança de vida nos grupos de idade e determinar quantidades de interesse em todas as faixas etárias no ano de 2010 na cidade de Ouro Preto-MG. Os dados foram coletados nos sites do IBGE e DATASUS. Foram aplicadas fórmulas específicas e obtidos os resultados e construída a tábua de mortalidade. Obtiveram-se os seguintes resultados: a faixa que ocorreu maior taxa de mortalidade é acima de 70 anos verificando-se 47 óbitos por mil habitantes e a faixa com maior esperança de vida é a de 0 a 4 anos com esperança de viver mais aproximadamente 78 anos. A probabilidade de morte de um indivíduo de idade exata  $x$  vir a morrer antes de completar  $x+n$  anos, por exemplo, um jovem de 15 anos morrer antes de completar 19 anos foi de 0,00629. Em Ouro Preto a esperança de vida da população com idade entre 15 anos e 19 anos era de aproximadamente 63 anos, três anos a mais do que esperado para o Brasil no ano de 2010.

**Palavras-chave:** *Tabua de Mortalidade, Esperança de vida, probabilidade.*

---

<sup>1</sup> Departamento de Estatística - Universidade Federal de Ouro Preto - UFOP, Ouro Preto-MG. Email: numiafgomes@yahoo.com.br

<sup>2</sup> Departamento de Estatística - Universidade Federal de Ouro Preto - UFOP, Ouro Preto-MG. Email: vanessa2miranda@yahoo.com.br

<sup>3</sup> Departamento de Estatística - Universidade Federal de Ouro Preto - UFOP, Ouro Preto-MG. Email: gragouvea@gmail.com

# Métodos de agrupamento em peixes redondos

Adriano C. Costa<sup>1,5</sup>

Tamara A. N. dos Anjos<sup>2,5</sup>

Marcelo A. Cirillo<sup>3,5</sup>

Rilke T. F. de Freitas<sup>4,5</sup>

**Resumo:** O objetivo deste trabalho foi avaliar métodos de ligação em análise de agrupamento na representação da distância genética de pacu, tambaqui, tambacu e paqui. Foram utilizados neste estudo 60 exemplares de pacu, 79 tambaqui, 43 tambacu e 43 paqui com 496 dias de idade do Setor de Piscicultura da Escola Agrotécnica Federal de Colatina. Os peixes foram pesados, submetidos à avaliação morfométrica e processados para obtenção dos pesos e dos rendimentos de corporais. Para comparar os métodos de ligação, ou seja, ligação centroide, completa, média, simples e ward, foi calculado o coeficiente de correlação cofenético, sendo utilizada a distancia Euclidiana dos fenótipos mensurados em todos os métodos. As análises foram realizadas no programa computacional R versão 3.0.1. Observou-se que todos os métodos de ligação apresentaram alta correlação cofenética e concordância nos resultados, sendo o pacu mas próximo do paqui e o tambaqui mais próximo do tambacu. O método de ligação simples foi o que apresentou menor valor de correlação cofenética (0.74), e o método de ligação pela média foi o que apresentou maior valor (0.79). Assim, o método de ligação pela media é o mais indicado para o agrupamento dos grupos genéticos estudados, pois apresentou menor grau de distorção que os demais métodos estudados.

**Palavras-chave:** *Colossoma macropomum*, *multivariada*, *Piaractus mesopotamicus*.

---

<sup>1</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [acarvalhocosta@gmail.com](mailto:acarvalhocosta@gmail.com)

<sup>2</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [ta\\_mara15@hotmail.com](mailto:ta_mara15@hotmail.com)

<sup>3</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [macufla@dex.ufla.br](mailto:macufla@dex.ufla.br)

<sup>4</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [rilke@dzo.ufla.br](mailto:rilke@dzo.ufla.br)

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

## Tamanho ideal da unidade experimental em ensaio com soja GMO Free

Guido Gustavo Humada-González<sup>1 3</sup>

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

Teodoro Adriano Bruzi<sup>3 3</sup>

Gilberto Rodrigues Liska<sup>4 3</sup>

Líder Ayala Aguilera<sup>5 3</sup>

Jose Rodolfo Humada Sosa<sup>6 3</sup>

Na agricultura a busca por aumento de produtividade cada vez mais se acelera com a constatação que em dia não tão distante não haverá mais terra para produzir e atender a necessidade mundial por alimentos. Uma saída do Brasil, líder mundial na produção e exportação de grãos GMO Free (livre de organismos geneticamente modificados), para satisfazer a crescente demanda da comunidade europeia por soja GMO Free é o desenvolvimento de variedades mais produtivas e executar experimentos para avalia-las. Para obter resultados confiáveis e de alta precisão é necessário utilizar experimentos bem planejados com tamanho e número de repetições adequados. Dado que não foram encontrados na literatura estudos recentes relacionados com tamanho de parcela em experimentos com soja convencional objetivou-se estimar o tamanho ótimo de parcela em soja GMO Free utilizando o método da curvatura máxima do coeficiente de variação. O experimento foi conduzido, de dezembro de 2013 a março de 2014, em DBC com três repetições na fazenda Experimental da Universidade Federal de Lavras, à 21° 14' latitude Sul, 45° 00' longitude Oeste. Cada unidade experimental básica da variedade BRS MG 752 S foi formada por 4 fileiras de 5 metros de comprimento, com distanciamento de 0,50 m entre fileiras. O tamanho de parcela é dado por:  $X_0 = 10\sqrt[3]{2(1-\hat{\rho}^2)}S^2\bar{Z}/\bar{Z}$  em que  $\hat{\rho}$  é a autocorrelação de primeira ordem,  $\bar{Z}$  é a média e  $S^2$  é a variância amostral. Concluiu-se que o tamanho ótimo estimado pelo método é 96 plantas equivalentes a 4 m<sup>2</sup>.

**Palavras-chave:** *Soja convencional, Métodos de curvatura, Precisão experimental.*

---

<sup>1</sup> DEX-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [gustavohumad@hotmail.com](mailto:gustavohumad@hotmail.com)

<sup>2</sup> DEX-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [armorais@dex.ufla.com.br](mailto:armorais@dex.ufla.com.br)

<sup>3</sup> DAG-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [adrianobruzi@dag.ufla.br](mailto:adrianobruzi@dag.ufla.br)

<sup>4</sup> DEX-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [gilbertoliska@hotmail.com](mailto:gilbertoliska@hotmail.com)

<sup>5</sup> FCA-UNA – Universidad Nacional de Asunción, San Lorenzo – PY Email: [semillas@agr.una.py](mailto:semillas@agr.una.py)

<sup>6</sup> FCA-UNA – Universidad Nacional de Asunción, San Lorenzo – PY Email: [jochumad@hotmail.com](mailto:jochumad@hotmail.com)

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

# Bootstrap e Cálculo de Tamanho amostral para Teste Não-Paramétrico de Postos Sinalizados de Wilcoxon

Natanael Silveira Mendes<sup>1</sup>

Sueli Aparecida Mingoti<sup>2</sup>

Em artigo publicado por Wang et al. (2003)<sup>3</sup> foram apresentados métodos para determinação do tamanho amostral a ser utilizado em diversos testes de hipóteses não-paramétricos. No caso específico do Teste de Postos Sinalizados de Wilcoxon, a fórmula exige que o pesquisador indique valores numéricos de determinados parâmetros ( $p_2, p_3, p_4$ ). Esses parâmetros possuem em sua formulação trechos em que são calculadas combinações que podem atingir ordem de grandeza elevadas, resultando em severo esforço computacional. Propõe-se, então, o uso de *bootstrap* e simulações Monte Carlo para estimação dos valores desses parâmetros por meio de um algoritmo implementado na linguagem do *software* estatístico R. A implementação é testada em um banco de dados real, resultando em convergência relativamente rápida das estimativas.

**Palavras-chave:** *Teste não-paramétrico, Postos Sinalizados de Wilcoxon, Bootstrap, Monte Carlo*

---

<sup>1</sup>Graduando em Estatística – UFMG Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [natanael.silveira.mendes@gmail.com](mailto:natanael.silveira.mendes@gmail.com)

<sup>2</sup>ICEX – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [sueli@est.ufmg](mailto:sueli@est.ufmg).

<sup>3</sup>Wang, H.; Chen B.; Chow, S.C. Sample size determination based on rank tests in clinical trials. Journal of Biopharmaceutical statistics, Vol. 13,4,pp-735-751. 2003

# Distribuições de Referência em Famílias de Distribuições com Núcleo Normal

Guilherme Lopes de Oliveira<sup>1 3</sup>

Rosângela Helena Loschi<sup>2 3</sup>

Elal-Olivero (2009), baseado na distribuição *Skew-Normal* (SN) introduzida por Azzalini (1985), derivou uma classe flexível de distribuições que acomodam formas uni e bimodais além da assimetria: a família de distribuições *Skew-Normal-Bimodal* (SNB). Neste trabalho apresentamos algumas questões relacionadas à inferência Bayesiana para esta família de distribuições. Mais especificamente, focamos no estudo da distribuição *a priori* não-informativa de Jeffreys para o parâmetro  $\alpha$  que controla a bimodalidade na família Normal-Bimodal (NB), estudando graficamente o comportamento desta distribuição de referência e derivando algumas de suas propriedades. Apresentamos também os resultados existentes na literatura com respeito à distribuição de Jeffreys para o parâmetro  $\lambda$  que controla a assimetria na família SN e estudamos ainda a distribuição de Jeffreys conjunta de  $\alpha$  e  $\lambda$  na família de modelos SNB considerando independência entre estes parâmetros.

**Palavras-chave:** *Skew-Normal-Bimodal, distribuição de Jeffreys, assimetria, bimodalidade.*

---

<sup>1</sup>UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [guilopes2110@gmail.com](mailto:guilopes2110@gmail.com)

<sup>2</sup>UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [loschi@est.ufmg.br](mailto:loschi@est.ufmg.br)

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

# Análise de Componentes Principais aplicado em Índices de Preços ao Consumidor

**João Flávio A. Silva**<sup>1</sup>

**Vitor Hugo Moreira**<sup>2</sup>

**Priscila Neves Faria**<sup>3</sup>

Os índices de preços foram criados com diversas finalidades como reajustar operações financeiras, reajustar salários, corrigir balanços e demonstrações financeiras e também como um medidor da inflação. As variações dos Índices de Preços ao Consumidor (INPC) são medidas ano a ano e, segundo várias categorias nas quais este índice é calculado, é possível analisar os dados históricos de forma a obter a redução da dimensão. Assim, as variações mensais das categorias Alimentos e Bebidas, Educação, Habitação, Artigos de Residência, Saúde, Vestuário, Transporte, Despesas Pessoais e Comunicação, ocorridas no período de janeiro de 2009 a dezembro de 2013, em 11 regiões Metropolitanas do Brasil, foram analisadas através da Análise de Componentes Principais (ACP). Esta análise é indicada quando o objetivo é a redução da dimensão, considerando uma perda aceitável de informação, retirando as variáveis que pouco contribuem para a variação total. Como resultado, os três primeiros componentes principais explicaram cerca de 76,3% da variação total. Além disso, foi observado que as variáveis que mais contribuíram para a variação total dos dados foram os Índices de Preços ao Consumidor de Educação, Alimentos e Bebidas e Despesas Pessoais. Já o Índice de Preço ao Consumidor de Comunicação foi a variável com menor contribuição, sendo esta variável a mais adequada a ser excluída do INPC. As análises estatísticas foram realizadas no software livre R.

**Palavras-chaves:** Análise Multivariada, Comparação, Economia.

---

<sup>1</sup> FAMAT - Universidade Federal de Uberlândia. Email: joao.1988@hotmail.com

<sup>2</sup> FAMAT - Universidade Federal de Uberlândia. Email: vitor-hugo-moreira@hotmail.com

<sup>3</sup> FAMAT - Universidade Federal de Uberlândia. Email: priscila@famat.ufu.br

# Distribuição de Frequências na análise do Clima Organizacional de uma Empresa Júnior

Jéssica Viviana Martins Mendes<sup>1,3</sup>

Priscila Neves Faria<sup>2,3</sup>

## 1-Introdução

As empresas precisam de funcionários satisfeitos, motivados, porem muitas vezes é difícil atingir essa motivação, pois o que motiva uns pode não motivar a outros. Fazendo com que as empresas acabem fracassando e se descuidando da importância das pessoas, colocando assim em risco o resultado a ser atingido pela organização.

Com esta preocupação, a pesquisa de Clima Organizacional (CO) vem se tornando cada vez mais importante devido ao fato de que busca detectar a qualidade do ambiente que é percebida ou experimentada pelos membros de uma empresa e que influencia o comportamento dos mesmos. Segundo [2], o CO é “aquela atmosfera psicológica que todas as pessoas percebem ao entrar num determinado ambiente e que as faz sentir mais ou menos à vontade para ali permanecer, interagir e realizar”.

O clima organizacional também é definido por [1] como “fenômeno perceptual duradouro, construído com base na experiência, multi-dimensional e compartilhado pelos membros de uma unidade da organização, cuja função principal é orientar e regular os comportamentos individuais de acordo com os padrões determinados por ela”.

Os instrumentos utilizados para avaliar o CO são as “pesquisas de clima”, ou seja, é uma pesquisa de opinião que caracteriza uma representação ou uma percepção da realidade organizacional.

Percebeu-se a importância de estudar o CO, a fim de melhorar a motivação e satisfação dos membros da empresa e conseqüentemente o contato dos mesmos com os líderes e diretorias. Sendo assim foram coletados dados que possibilitaram um reconhecimento da real situação do ambiente de trabalho da empresa.

---

<sup>1</sup>Faculdade de Matemática - FAMAT / UFU- Universidade Federal de Uberlândia ,Uberlândia -MG.Email : [jessicaviviana@est.ufu.br](mailto:jessicaviviana@est.ufu.br);

<sup>2</sup> Faculdade de Matemática - FAMAT / UFU- Universidade Federal de Uberlândia ,Uberlândia -MG .Email: [priscila@famat.ufu.br](mailto:priscila@famat.ufu.br);

<sup>3</sup> Agradecimento a FAPEMIG pelo apoio financeiro

## **2- Objetivo**

Este trabalho tem o objetivo de apresentar um estudo sobre CO a fim de conhecer a frequência absoluta (número de ocorrências em cada categoria), a frequência relativa (frequência absoluta representada em porcentagem referente ao total), satisfação e detecção de soluções.

## **3-Metodologia**

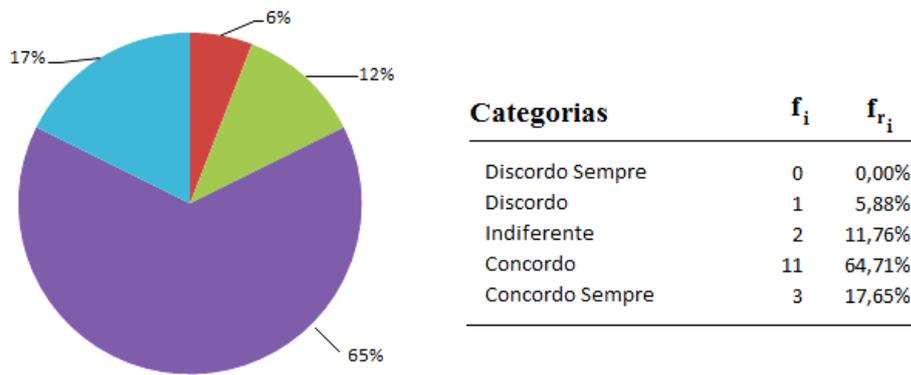
Para esta pesquisa, tomou-se base o questionário do modelo Great Place to Work que avalia cinco fatores: credibilidade, respeito, imparcialidade, orgulho e camaradagem. Para garantir o sigilo das respostas individuais, o questionário não possui identificação. O mesmo, tem como campos de respostas uma escala do tipo Likert com cinco níveis de avaliação: 1 – discordo sempre; 2 – discordo na maioria das vezes; 3 – às vezes concordo, às vezes discordo; 4 – concordo na maioria das vezes; 5 – concordo sempre. Assim, a presente pesquisa caracteriza-se como quantitativa e qualitativa.

Foi realizada uma pesquisa de campo e através de um questionário conforme a escala *Likert*, foram abrangidas variáveis que identificaram os fatores que influenciam o clima organizacional da empresa. O questionário foi aplicado a 17 membros, ou seja, em 80,95% dos membros ativos.

O questionário é o instrumento de coleta da pesquisa quantitativa e ele deve ser adequado aos objetivos do estudo, ao público que está sendo pesquisado, à técnica de coleta dos dados e ao tempo de aplicação. Existem muitos segredos que fazem de um questionário um eficiente instrumento de coleta e o cuidado com as escalas é um deles. Assim, as escalas permitem coletar diferentes informações de forma a avaliar, graduar, comparar, diferenciar, ordenar as opiniões registradas em cada abordagem do questionário.

## **4-Resultados e Discussão**

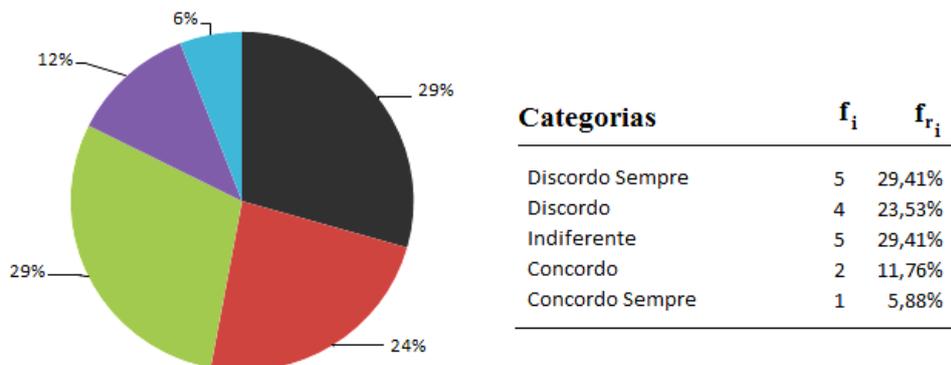
Quanto à pergunta “A CONSELT é um lugar agradável para trabalhar?”, observou-se que a imagem da empresa perante seus membros é positiva visto que a maioria das respostas foram relacionadas à categoria ‘Concordo’ (cerca de 65%). Este resultado ressalta que os entrevistados possuem uma positiva visualização da CONSELT e do inter-relacionamento existente entre membros e diretorias que a compõem.



**Figura 1-** Categorização da satisfação quanto ao ambiente de trabalho.

Vale ressaltar que a categoria ‘Indiferente’ não é considerada aqui como negativa e nem como positiva uma vez que, se um entrevistado não possui opinião formada a respeito, isto talvez indique que o mesmo pode ter entrado na empresa recentemente.

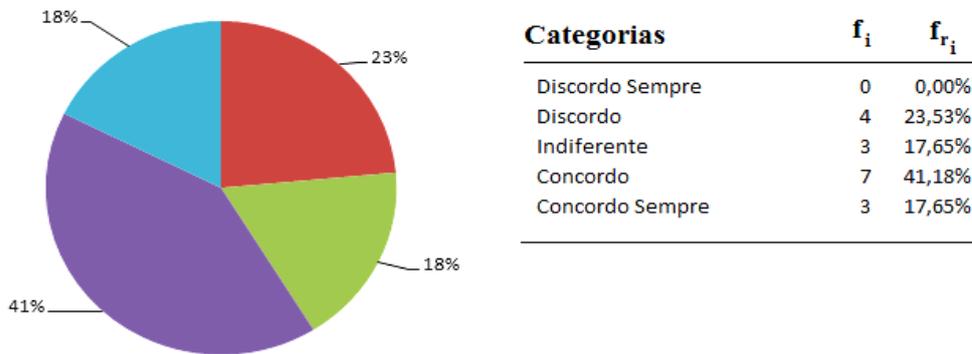
Na Figura 2, observa-se o resultado obtido referente à pergunta “As pessoas são recompensadas adequadamente pelo serviço que fazem?”, havendo um empate nas respostas do tipo ‘Discordo Sempre’ e ‘Indiferente’, sendo que 23,53% registram a opção ‘Discordo’. Este resultado foi extremamente negativo, indicando a insatisfação dos membros da empresa quanto ao conquistado a cada meta e trabalho realizado por eles. Isto sugere que a recompensa pelos trabalhos prestados deve ser implantada, podendo esta ser no custeio de capacitação e integração dos seus membros, por não ser dos ideais da empresa oferecer recompensas financeiras.



**Figura 2** – Categorização da satisfação quanto à recompensa que recebem referente aos trabalhos prestados.

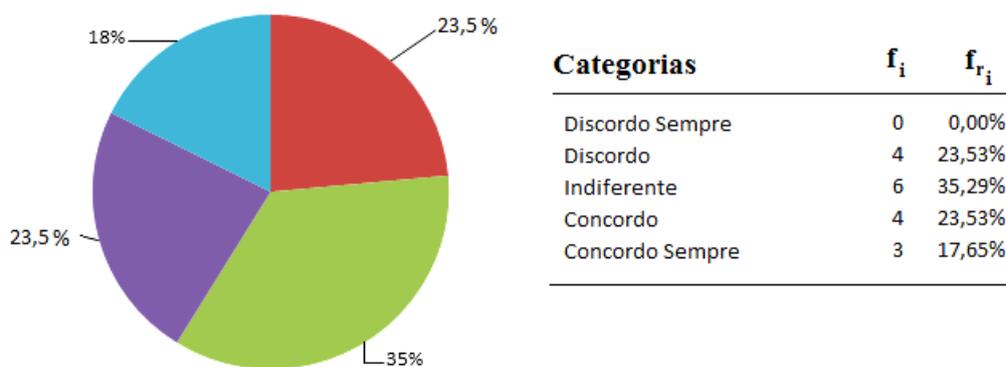
A Figura 3 mostra o resultado baseado na pergunta “Os líderes estão interessados nas ideias e sugestões que os demais membros dão, existindo tomada de medidas com base nisto?” Houve predominância na resposta ‘Concordo’ (cerca de 41%), o que demonstra opinião positiva dos membros da CONSELT em relação à pergunta feita. Desta forma, observa-se que existe a

preocupação com a coleta de novas ideias e sugestões por parte de todos os setores da empresa, havendo tomada de decisão e execução das ideias, com base nas atividades exercidas pela mesma.



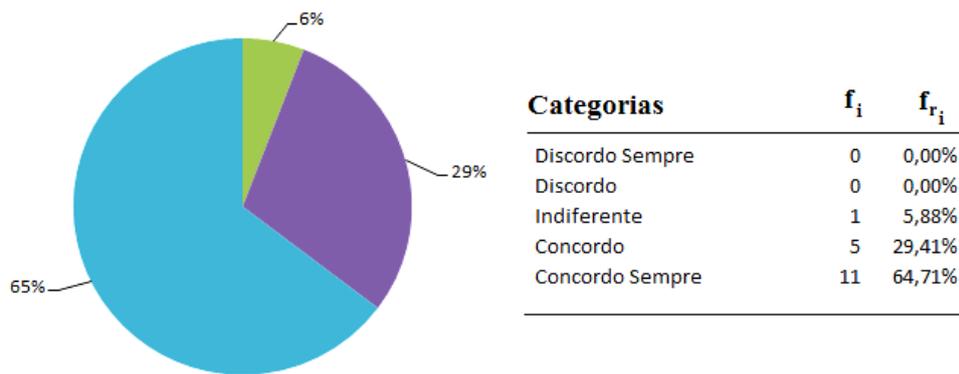
**Figura 3** – Categorização da satisfação quanto à preocupação da empresa na coleta de novas ideias e sugestões.

A Figura 4 mostra o resultado quanto à pergunta “Os líderes envolvem seus membros em decisões que afetam suas atividades e seu ambiente de trabalho?” sendo que os entrevistados (membros da empresa) avaliaram, em sua maioria, de forma ‘Indiferente’ (cerca e 35%). Além disso, houve empate quanto às respostas opostas ‘Discordo’ e ‘Concordo’ (23,5% aproximadamente). Este resultado pode auxiliar a empresa na execução de ideias, de forma que, na medida do possível, sejam englobadas as opiniões dos membros antes que decisões sejam tomadas.



**Figura 4** - Gráfico de setores referente a opinião dos entrevistados quanto à postura da empresa na tomada de decisões.

Na Figura 7 observa-se o resultado quanto ao questionamento “Tenho orgulho de contar à outras pessoas que trabalho aqui”. A maioria respondeu positivamente (cerca de 94%) quanto à satisfação em dizer que trabalha para a empresa, demonstrando um certo orgulho em pertencer ao grupo, fundamentado também na pergunta 5 (Figura 5), sendo observado que, tanto a participação dos membros na empresa quanto à coleta dos resultados por parte deles, é satisfatória.



**Figura 7** - Gráfico referente ao questionamento quanto ao orgulho em pertencer ao grupo.

### 5-Conclusão

Desta forma, conclui-se que os membros da empresa CONSELT tem uma visão positiva da mesma, bem como do trabalho que realizam junto à ela. No entanto, é importante a busca de resultados ainda melhores visando à plena satisfação da equipe na busca da consolidação de opiniões que tangem o clima organizacional da empresa.

### Referências

- [1] Koys, D., & DeCotiis, T. (1991) Inductive measures of psychological climate. *Human Relations*, 44, (3), 265-285.
- [2] MALLETT, R. *O Poder Do Ambiente Sobre A Produtividade*, 2009. Disponível em: <http://www.artigonal.com/negocios-admin-artigos/clima-organizacional-o-poder-do-ambiente->

# Comparação de atributos sensoriais de hamburger via média ponderada ordinal

Marcos Costa de Paula<sup>1 3</sup>

Mateus Pimenta Siqueira Lima<sup>2</sup>

Eric Batista Ferreira<sup>3 3</sup>

Hambúrguer é um alimento feito a base de carne (geralmente bovina ou de frango), muito utilizada na alimentação ocidental em lanches rápidos (a chamada *fast food*). Na decisão de compra por uma marca ou outra, o consumidor leva em consideração o preço, a marca, a embalagem e, é claro, a qualidade intrínseca do produto. Essa, por sua vez, pode ser didaticamente dividida em qualidade microbiológica ou sanitária, físico-química e sensorial. Na avaliação sensorial de alimentos, grupos de provadores chamados painéis são treinados para avaliar e comprar características tais como sabor característico, cor, textura, aparência, etc, variando de produto para produto. Um experimento foi realizado na Universidade Federal de Alfenas para comparar 5 marcas comerciais de hambúrguer por meio de um conjunto de atributos sensoriais gerado pelo próprio painel sensorial. Foi utilizada a escala hedônica de nove pontos para pontuá-los e, portanto, foi utilizada uma métrica capaz de lidar com postos, a média ponderada ordinal (MPO). O resultados mostram a adequação dessa métrica à comparação de marcas de hambúrguer, por as variáveis resposta estritamente como são, categóricas.

**Palavras-chave:** *Média ponderada ordinal, hambúrguer, sensorial.*

---

<sup>1</sup>Graduando em Matemática. Universidade Federal de Alfenas. Email: *costa\_mvp@hotmail.com*

<sup>2</sup>Graduando em Matemática. Universidade Federal de Alfenas. Email: *mateuspimenta-2007@hotmail.com*

<sup>3</sup>Instituto de Ciências Exatas. Universidade Federal de Alfenas. Email: *eric.ferreira@unifal-mg.edu.br*

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

# Aplicação de análise de sobrevivência aos dados de hanseníase da cidade de Patos de Minas - MG

Bruno Henrique Gomes<sup>1 3</sup>

Fabírcia de Matos Oliveira<sup>2</sup>

A Análise de Sobrevivência é um ramo da estatística, aplicada principalmente a dados epidemiológicos, quando a variável em estudo representa o tempo desde um instante inicial bem definido até à ocorrência de determinado acontecimento de interesse. Este estudo teve como objetivo aplicar análise de sobrevivência aos fatores de risco associados ao tempo, até a cura de pacientes com hanseníase, diagnosticados no período de janeiro de 2003 a dezembro de 2013 na cidade de Patos de Minas - MG. Foram utilizados dados, fornecidos pela Superintendência Regional de Saúde, de 540 pacientes com a doença. As funções de sobrevivência foram estimadas por métodos não paramétricos por meio do Estimador de Kaplan-Meier, as decisões estatísticas foram tomadas ao nível de significância de 0,1. Os resultados sugerem que pacientes com mais de 40 anos tem a probabilidade de cura de 57,12% a mais do que pacientes com idade menor que 40 anos; pacientes com classificação operacional paucibacilar tem 40,9% a mais na probabilidade de cura do que pacientes de classificação operacional multibacilar; pacientes com ensino superior completo ou incompleto tem a probabilidade de cura de 26,43% a mais do que pacientes com menor grau de escolaridade.

**Palavras chaves:** *Sobrevivência, Kaplan-Meier, Métodos não paramétricos.*

---

<sup>1</sup>INGEB-UFU: Instituto de Genética e Bioquímica - Universidade Federal de Uberlândia, Patos de Minas-MG. Email: [b.hgomes@hotmail.com](mailto:b.hgomes@hotmail.com)

<sup>2</sup>FAMAT-UFU: Faculdade de Matemática - Universidade Federal de Uberlândia, Patos de Minas-MG. Email: [fabriciamatos@famat.ufu.br](mailto:fabriciamatos@famat.ufu.br)

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

# ESTIMAÇÃO DOS RETORNOS DAS AÇÕES DE COMPANHIAS DE SEGURO NO BRASIL: UMA ABORDAGEM BAYESIANA DO MODELO CAPM

Paulo Henrique Medeiros Bragança<sup>15</sup>

Marçal Serafim Cândido<sup>25</sup>

Leandro Ferreira<sup>35</sup>

Luiz Carlos Funchal Junior<sup>45</sup>

No atual contexto econômico, muito são os agentes, físicos e jurídicos que buscam investir em títulos de renda variável, especialmente em ações. Com isso o mercado de capitais, forjado pelas diversas bolsas de valores existentes, fica cada vez mais dinâmico, e a variável que, cada vez mais, redireciona a atenção dos especialistas da área financeira é o risco. Desta forma, procurou-se neste trabalho, abordar o uso da inferência bayesiana no modelo CAPM, que é um modelo de precificação de ativos em bolsa de valores, possibilitando encontrar retornos esperados para um ativo qualquer como também encontrar o risco, ou volatilidade, de tal ativo no mercado. Dentro de tal metodologia, foi adicionado o índice de sinistralidade no modelo, buscando entender se este índice influencia as decisões dos agentes investidores nas ações de empresas seguradoras. De acordo com o modelo proposto, não houve a existência de influências externas sobre os ativos considerados no trabalho, além do movimento de cada ativo estar de acordo com o mercado. Pode-se também observar que a sinistralidade não foi significativa no modelo proposto.

**Palavras-chave:** *CAPM, inferência bayesiana, análise de risco.*

<sup>1</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [braganca\\_paulo@live.com](mailto:braganca_paulo@live.com)

<sup>2</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [marcal.serafim@unifal-mg.edu.br](mailto:marcal.serafim@unifal-mg.edu.br)

<sup>3</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [leandro.ferreira@unifal-mg.edu.br](mailto:leandro.ferreira@unifal-mg.edu.br)

<sup>4</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [luz\\_funchal@hotmail.com](mailto:luz_funchal@hotmail.com)

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

# Influência dos Erros de Classificação no Desempenho do Planejamento Econômico para Controle On-line por Atributo com Amostras Não-unitárias

Lupércio França Bessegato<sup>1</sup>

Augusto dos Reis Pereira<sup>2</sup>

Roberto da Costa Quinino<sup>3</sup>

O modelo de Taguchi para controle on-line por atributo com amostra unitária consiste em amostrar um único item a cada  $m$  itens produzidos. Em cada inspeção, decide-se sobre o estado do processo de produção. Caso o item inspecionado seja declarado não conforme, supõe-se que o processo esteja fora de controle, ajustando-o imediatamente para reconduzi-lo a sua condição inicial. Para planejar esse modelo é utilizada uma estratégia econômica, cujo objetivo é determinar o intervalo ótimo de inspeção que minimize o custo esperado por item em um ciclo de produção. Este trabalho apresenta uma extensão ao modelo de Taguchi, em que são utilizadas amostras não unitárias. Considera-se também um sistema de classificação imperfeito, ou seja, as inspeções estão sujeitas a erros de classificação. A modelagem probabilística do sistema de controle utiliza as propriedades de cadeias de Markov ergódicas, obtendo-se a expressão do custo médio do sistema por item produzido. Esse custo é minimizado por três parâmetros: o intervalo entre inspeções,  $m$ , o tamanho da amostra,  $n$  e o número mínimo de classificações conformes para se declarar o processo sob controle,  $a$ . É utilizada uma abordagem por algoritmos genéticos a fim de encontrar os valores ótimos dos parâmetros de planejamento do sistema de controle. O procedimento é ilustrado por um exemplo numérico e é realizada uma análise de sensibilidade para estudar a influência de alguns parâmetros probabilísticos e de custo no desempenho do modelo.

**Palavras-chave:** controle on-line por atributo, erros de classificação, amostras não-unitárias, cadeia de markov, planejamento econômico.

---

<sup>1</sup> UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. E-mail: [lupercio.bessegato@ufjf.edu.br](mailto:lupercio.bessegato@ufjf.edu.br)

<sup>2</sup> UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. E-mail: [augusto.pereira@ice.ufjf.br](mailto:augusto.pereira@ice.ufjf.br)

<sup>3</sup> UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. E-mail: [roberto@est.ufmg.br](mailto:roberto@est.ufmg.br)

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

# Estudo de caso para detecção prospectiva de conglomerados espaço-tempo em microblogs: o Carnaval no Brasil

Thais Rotsen Correa<sup>1</sup>

Bráulio Miranda Veloso<sup>2</sup>

Andrea Iabrudi Tavares<sup>3</sup>

Neste trabalho utilizamos um sistema de vigilância para identificar conglomerados emergentes em tempo real, a partir de evidências fornecidas pela informação espaço-temporal de um microblog durante o Carnaval de 2012 no estado brasileiro de São Paulo. Este tipo de abordagem vem sendo utilizada com sucesso em várias aplicações, como acompanhamento de epidemias, catástrofes e eventos sociais. Adotamos um sistema adequado às peculiaridades das postagens no Twitter, capaz de localizar precocemente concentrações anormais de Tweets que podem ser evidência de acontecimentos no mundo real. Este sistema - que é uma variação do método proposto por Assunção e Correa (2009) - utiliza uma janela temporal que, além de diminuir o tempo de execução, reflete o interesse em acontecimentos recentes. O método adotado apresentou resultados satisfatórios quando comparado à estatística scan, que é uma das principais técnicas para detecção de conglomerados.

**Palavras-chave:** *Alarme, Conglomerado espaço-tempo, Janela temporal, Sistema de vigilância.*

---

<sup>1</sup>DEST-UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [thais@est.ufmg.br](mailto:thais@est.ufmg.br)

<sup>2</sup>DCC-UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [bmveloso@dcc.ufmg.br](mailto:bmveloso@dcc.ufmg.br)

<sup>3</sup>Jasper Design Automotivation Brasl, Belo Horizonte-MG. Email: [andrea.iabrudi@gmail.com](mailto:andrea.iabrudi@gmail.com)

# UM ESTUDO DA DINÂMICA POPULACIONAL DO MUNICÍPIO DE VARGINHA-MG UTILIZANDO LÓGICA FUZZY

Luiz Carlos Funchal Junior<sup>1 3</sup>

Leandro Ferreira<sup>2 3</sup>

Com a finalidade de prever o comportamento populacional de municípios brasileiros, estudos têm sido realizados utilizando modelos determinísticos e modelos baseados na lógica *fuzzy*. No presente trabalho, um estudo foi realizado para analisar a dinâmica populacional do município de Varginha-MG, através da utilização de um modelo baseado em regras *fuzzy* (*fuzziness* demográfica). Para tanto, foram analisados os erros obtidos entre dados coletados via IBGE e obtidos pelo modelo determinístico de Verhulst. De acordo com os resultados obtidos, modelos baseados na lógica *fuzzy* podem ser utilizados de maneira satisfatória para o estudo da dinâmica populacional do município de Varginha-MG.

**Palavras-chave:** *dinâmica de populações, fuzziness demográfica, modelos determinísticos.*

<sup>1</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [luiz\\_funchal@hotmail.com](mailto:luiz_funchal@hotmail.com)

<sup>2</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [leandro.ferreira@unifal-mg.edu.br](mailto:leandro.ferreira@unifal-mg.edu.br)

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

# A Distribuição Beta Fréchet Transmutada

Marcio Claudio Mercês Brito<sup>1</sup>

Jailson de Araujo Rodrigues<sup>2</sup>

Jaime dos Santos Filho<sup>3</sup>

Ana Paula Coelho Madeira Silva<sup>4</sup>

Guido Gustavo Humada González<sup>5</sup>

A distribuição Fréchet ou Weibull inversa como também é conhecida foi originalmente proposta para analisar dados de confiabilidade e para investigar falhas de componentes mecânicos sujeitos a degradação. Com o decorrer dos anos, a distribuição Fréchet passou a ser um dos modelos mais utilizados na análise de dados de sobrevivência, isso levou ao surgimento de generalizações e extensões dessa distribuição. O objetivo deste trabalho é introduzir e estudar uma distribuição com cinco parâmetros denominada beta Fréchet transmutada. São deduzidas algumas das principais propriedades estruturais dessa distribuição. Em particular, são apresentadas expressões para os momentos e a função geradora de momentos. O ajuste do modelo a um conjunto de dados demonstra que o mesmo pode ser utilizado como alternativa no ajuste de dados de sobrevivência.

**Palavras-chave:** *Comparação de Modelos, Distribuição Fréchet, Estimadores de Máxima Verossimilhança.*

---

<sup>1</sup>Instituto Federal Baiano-IFBaiano. Email: [marcinhofab@hotmail.com](mailto:marcinhofab@hotmail.com)

<sup>2</sup>Instituto Federal da Bahia-IFBA. Email: [jailsondearaujo@yahoo.com.br](mailto:jailsondearaujo@yahoo.com.br)

<sup>3</sup>Instituto Federal da Bahia-IFBA. Email: [jst\\_ba@yahoo.com.br](mailto:jst_ba@yahoo.com.br)

<sup>4</sup>Universidade Federal de São João Del Rei-UFSJ. Email: [apcmadeira@hotmail.com](mailto:apcmadeira@hotmail.com)

<sup>5</sup>Universidade Federal de Lavras- UFLA. Email: [gustavohumad@hotmail.com](mailto:gustavohumad@hotmail.com)

# Correlação canônica entre variáveis morfométricas e peso das partes corporais em Pacu

Adriano Carvalho Costa<sup>1 5</sup>

Érica F. da Cruz<sup>2 5</sup>

Marcelo A. Cirillo<sup>3 5</sup>

Rilke T. F. de Freitas<sup>4 5</sup>

**Resumo:** A análise de correlação canônica tem como objetivo principal o estudo das relações lineares existentes entre dois conjuntos de variáveis. O objetivo deste trabalho foi verificar se há uma combinação linear entre um grupo de variáveis morfométricas e um grupo de variáveis composto pelos pesos corporal, carcaça e do filé de pacu, que maximize a correlação entre os dois grupos. Foram utilizados neste estudo 162 exemplares de pacu, com peso médio de 872g ( $\pm 164$ g), retirados aleatoriamente de uma população de 500 peixes do plantel de animais do Setor de Piscicultura da Escola Agrotécnica Federal de Colatina. Os peixes foram submetidos à avaliação morfométrica e processados para obtenção dos pesos da carcaça e do filé. A correlação canônica entre os grupos de variáveis foi realizada utilizando o programa computacional R. Observou-se correlação alta e significativa ( $r=0,98$ ) para o primeiro par canônico. Dessa forma, o primeiro par canônico  $U1 = -0,00319PESO - 0,00378CARCAÇA + 0,00101FILE$  e  $V1 = -0,31194CP + 4,08925CC - 2,54718AD + 0,70299LD - 179,75724CC/CP + 84,86118AD/CP - 37,33908LD/CP + 2,39401CC/AD - 20,86389LD/CC + 50,80873LD/AD$  pode ser usado para representação dos grupos estudados. Pode-se concluir através deste trabalho, que há uma combinação linear entre o grupo de variáveis morfométricas e o grupo de pesos das partes corporais em pacu.

**Palavras-chave:** *associação, Piactus mesopotamicus, peixes*

---

<sup>1</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [acarvalhocosta@gmail.com](mailto:acarvalhocosta@gmail.com)

<sup>2</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [erica.89@bol.com.br](mailto:erica.89@bol.com.br)

<sup>3</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [macufla@dex.ufla.br](mailto:macufla@dex.ufla.br)

<sup>4</sup> UFLA: Universidade Federal de Lavras, Lavras – MG. Email: [rilke@dzo.ufla.br](mailto:rilke@dzo.ufla.br)

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

# Uso de séries temporais em um estudo do Índice Pluviométrico no município de Flores – PE

Danielle Gonçalves de Oliveira Prado<sup>1 6</sup>

Larissa Ribeiro de Andrade<sup>2 6</sup>

Victor Moreira Gonçalves<sup>3 6</sup>

Rodrigo de Andrade Ferrazza<sup>4 6</sup>

Isabelle Gonçalves de Oliveira Prado<sup>5 6</sup>

O estudo do índice pluviométrico torna-se essencial na tentativa de reduzir os impactos causados pela insuficiência de precipitação pluviométrica em determinada região, através de previsões. Com o objetivo de ilustrar o método de análise de séries temporais, foi utilizado um conjunto de dados oriundos de coletas mensais desde jan/1985 até dez/2000 em uma estação meteorológica localizada em Flores, região brasileira que sofre com a seca, localizada na porção norte de Pernambuco. Inicialmente, foi analisado o gráfico da amplitude versus a média, de forma a verificar se era necessário aplicar transformação logarítmica nos dados. Posteriormente, para verificar a presença de tendência e sazonalidade, foi feita a análise da função de autocorrelação e função de autocorrelação parcial, além de serem utilizados o teste do sinal (tendência) e o teste de Fisher (sazonalidade), descritos em Morettin e Tolo (2006). Foram sugeridos os modelos  $SARIMA(2,1,1)(0,1,1)_{12}$  e  $SARIMA(1,1,1)(0,1,1)_{12}$  que apresentaram AIC próximos. Pela parcimônia, considerou-se o modelo  $SARIMA(1,1,1)(0,1,1)_{12}$  para fornecer as previsões. A análise de resíduo foi feita utilizando o teste de Box-Pierce (MORETTIN & TOLOI, 2006). Baseado na previsão realizada, os meses mais críticos no ano de 2001 seriam os meses de agosto, setembro e outubro. Seria interessante, portanto, realizar campanhas que alertassem os moradores da região para esse fato e incentivar campanhas políticas que os abasteçam nesse período.

**Palavras-chave:** *Séries temporais, seca, insuficiência pluviométrica.*

---

<sup>1</sup> DEX – UFLA e-mail: danigoprado@hotmail.com

<sup>2</sup> DEX – UFLA e-mail: larissa.mat@gmail.com

<sup>3</sup> FAMAT – UFU e-mail: victormoreira\_28@hotmail.com

<sup>4</sup> FMVZ – UNESP/Botucatu. e-mail: rodrigoferrazza@yahoo.com.br

<sup>5</sup> DBI – UFV e-mail: bellegoprado@hotmail.com

<sup>6</sup> Agradecimento à FAPEMIG pelo apoio financeiro.

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Estatística

Distribuição Weibull:  
uma aplicação aos modelos de  
degradação

Aluna: Jéssica da Assunção Almeida  
Orientador: Fábio Nogueira Demarqui

Belo Horizonte - MG, dezembro de 2013

## Resumo

O acesso à confiabilidade de dispositivos é, tradicionalmente, baseado em testes de vida acelerados. No entanto, em uma gama de situações envolvendo produtos altamente confiáveis, pouca informação acerca da confiabilidade destes produtos pode ser acessada, visto que um número significativamente reduzido de falhas é observado durante o período em que o produto permanece sob teste. Em situações nas quais é razoável a suposição de que as falhas são induzidas por um mecanismo de degradação, é possível acessar a confiabilidade de dispositivos através da informação proveniente do seu mecanismo de degradação, obtida após um período de monitoramento dos dispositivos sob investigação. Neste trabalho um conjunto de dados real foi analisado com o intuito de ilustrar como dados de degradação podem ser utilizados para se acessar a confiabilidade de produtos. Os dados de degradação foram modelados através do método analítico e o método aproximado. Para ambos os métodos, a distribuição Weibull foi assumida para modelar a distribuição do tempo até a falha, e abordagens frequentista e Bayesiana foram consideradas. O ajuste dos diferentes modelos foi realizado através dos softwares estatísticos R e OpenBugs. Com base nos modelos ajustados foram obtidas medidas importantes no estudo de confiabilidade, como a função de confiabilidade e percentis da distribuição do tempo de vida dos produtos. Finalmente, foi observado que, para o conjunto de dados analisado, os métodos analítico e aproximado apresentaram resultados similares, independentemente da abordagem (frequentista ou Bayesiana) adotada. Este trabalho é desenvolvido com auxílio financeiro da FAPEMIG.

**Palavras-chave:** Confiabilidade, Inferência Bayesiana, Método Analítico, Método Aproximado, Métodos MCMC

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Confiabilidade</b>	<b>3</b>
<b>3</b>	<b>Inferência Bayesiana</b>	<b>6</b>
<b>4</b>	<b>Modelos de degradação</b>	<b>9</b>
4.1	Método Analítico . . . . .	9
4.2	Método Aproximado . . . . .	12
<b>5</b>	<b>Estudo de caso</b>	<b>13</b>
<b>6</b>	<b>Considerações Finais</b>	<b>18</b>
<b>7</b>	<b>Apêndice</b>	<b>19</b>
7.1	Gráficos de Diagnóstico . . . . .	19
7.2	Código em <i>OpenBUGS</i> . . . . .	20
<b>8</b>	<b>Referências Bibliográficas</b>	<b>23</b>

# 1 Introdução

O acesso à confiabilidade de dispositivos é, tradicionalmente, baseado em testes de vida acelerados. No entanto, em uma gama de situações envolvendo produtos altamente confiáveis, pouca informação acerca da confiabilidade destes produtos pode ser acessada, visto que um número significativamente reduzido de falhas é observado durante o período em que o produto permanece sob teste. Neste contexto, a análise de degradação surge como uma alternativa eficiente quando é possível supor que a falha do dispositivo em estudo é dada por um mecanismo de degradação.

Na literatura, vem sendo mostrado que os dados de degradação são uma alternativa superior, quando se tem pouca informação sobre o tempo de vida. Neste trabalho replicaremos o estudo de Hamada (2005), objetivando ilustrar como dados de degradação podem ser modelados e analisados.

Os dados apresentados em Hamada (2005) são referentes a degradação de laser e pertencem a um estudo realizado por Meeker e Escobar (1998). Numa corrente de operação fixa, a saída de luz de um laser se degrada ao longo do tempo. É preciso que essa saída seja mantida constante. Desta forma, a falha do laser é definida quando a corrente atinge 10%. Quinze unidades foram medidas de 0 a 4000 horas, a cada 250 horas.

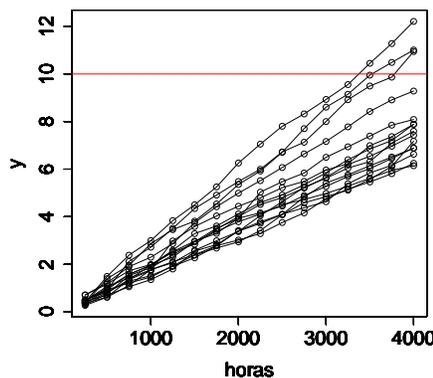


Figura 1: Gráfico dos dados de degradação de laser

Na Figura 1 é observado que a degradação de cada laser possui comportamento linear sem intercepto, pois no tempo zero a degradação de todos os laser também é zero.

Apresentamos modelagens através do método analítico e o método aproximado considerando abordagens Bayesiana e frequentista. Como a distribuição *a posteriori* não pode ser calculada numericamente, os softwares estatísticos R e OpenBUGS foram utilizados para geração das amostras *a posteriori*, através de métodos MCMC, e análise de resultados.

Este trabalho está organizado da seguinte forma. Na Seção 2 apresentamos uma breve descrição de confiabilidade. A Seção 3 é destinada a discussão de alguns conceitos utilizados em inferência Bayesiana. A metodologia geral para o ajuste de modelos de degradação é apresentada na Seção 4. Na Seção 5 uma replicação de uma análise de dados reais é apresentada. Finalmente, na Seção 6 são apresentadas as considerações finais sobre o trabalho.

## 2 Confiabilidade

A confiabilidade é um ramo da estatística que atua, principalmente, com falhas de componentes ou sistemas industriais, possuindo grande importância nas engenharias, mas podendo ser aplicada às mais diversas áreas de conhecimento.

A intensa competição no mercado aliada ao desenvolvimento tecnológico, pressiona a produção de produtos de alta qualidade. Por outro lado, consumidores esperam que sistemas, veículos, máquinas, dispositivos e etc, sejam confiáveis e seguros, ou seja, com alta probabilidade, sejam capazes de desempenharem suas funções pretendidas, por algum período de tempo especificado. Este contexto resultou no desenvolvimento de métodos probabilísticos e estatísticos para o tratamento das questões industriais relacionadas à confiabilidade.

Formalmente, Hamada (2008) define confiabilidade como a probabilidade de um produto realizar sua função pretendida para seu tempo de vida especificado, sob condições de operações indicadas, de um modo que atende ou excede as expectativas dos clientes.

Os estudos em confiabilidade envolvem o tempo até a ocorrência de um evento de interesse, sendo na maioria das vezes definido como falha. Porém, os testes são realizados envolvendo produtos altamente duráveis, portanto, finalizados antes que todos os itens falhem, resultando a ocorrência de informações censuradas, que por sua vez trazem pouca informação acerca da confiabilidade, de acordo com Hamada (2005).

Meecker (1998) descreve que a maioria dos dados de confiabilidade podem ser modelados utilizando distribuições para variáveis aleatórias positivas como as distribuições Weibull e log-normal.

Em situações nas quais é razoável a suposição de que as falhas são induzidas por um mecanismo de degradação, é possível acessar a confiabilidade de dispositivos através da informação proveniente do seu mecanismo de degradação, obtida após um período de monitoramento dos dispositivos sob investigação.

Segundo Freitas (2009), a análise de dados de degradação pode ser uma alternativa superior a dados altamente censurados e que dependerá da quantidade de erro com o qual o mecanismo é medido. Essa análise é baseada na escolha de uma característica degradante, que será relacionada ao fracasso. As falhas definidas em termos dessa característica são chamadas de falha suave, ou seja, quando um dispositivo atingir um nível crítico de degradação.

Assim como os dados de confiabilidade, os dados de degradação podem ser usados para modelar o tempo até a falha através de distribuições de probabilidade, das quais também se destacam a Weibull e a log-normal. Uma vez obtida a distribuição, medidas importantes no estudo da confiabilidade podem ser obtidas, como a função de confiabilidade, função

risco e percentis da distribuição do tempos de vida.

A distribuição Weibull é popular em aplicações práticas devido a variedade de formas, tendo como propriedade básica a função risco mónotona.

A função densidade de probabilidade é dada por:

$$f(t) = \beta\lambda t^{\beta-1} \exp\{-\lambda t^\beta\}, t \geq 0,$$

em que  $\gamma$ , o parâmetro de forma, e  $\alpha$ , o de escala, são ambos positivos. Para  $\gamma = 1$ , a distribuição exponencial é um caso particular da Weibull.

As funções de confiabilidade e risco, e o percentil  $\alpha$  são, respectivamente:

$$P(T > t) = R(t) = \exp\{-\lambda t^\beta\},$$

$$\lambda(t) = \beta\lambda t^{\beta-1}$$

e

$$t_\alpha = [-1/\lambda \ln(1 - \alpha)]^{1/\beta}$$

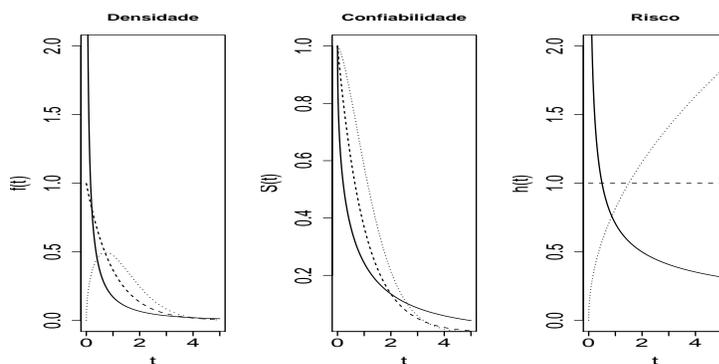


Figura 2: Forma típica das funções de densidade de probabilidade, confiabilidade e risco considerando  $\alpha = \gamma = 0,5(-); 1,0(---)$  e  $1.5(\dots)$ , da distribuição Weibull.

A distribuição Log-normal, assim como a Weibull, é bastante popular, sendo que os parâmetros  $\mu$  é a média do logaritmo do tempo de falha, assim como  $\sigma$  é o desvio padrão. Diferente da distribuição Weibull, as funções risco da distribuição log-normal não são monótonas. Segundo Colosimo (2006) elas crescem, atingem um valor máximo e depois decrescem. A função densidade de probabilidade é dada por:

$$f(t) = \frac{1}{\sqrt{2\pi t\sigma}} \exp \left\{ -\frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right\}, t \geq 0$$

As funções de confiabilidade e risco de uma variável log-normal não apresentam uma forma analítica explícita e é, desse modo, representadas, respectivamente, por:

$$P(T > t) = R(t) = \Phi \left( \frac{-\log(t) + \mu}{\sigma} \right),$$

e

$$\lambda(t) = \frac{f(t)}{S(t)}$$

onde  $\Phi$  é a distribuição acumulada da normal padrão.

Os percentis  $\alpha$  para a distribuição log-normal são obtidos através da seguinte expressão:

$$t_\alpha = \exp z_\alpha \sigma + \mu$$

em que  $z_\alpha$  é o  $100\alpha\%$  percentil da distribuição normal padrão.

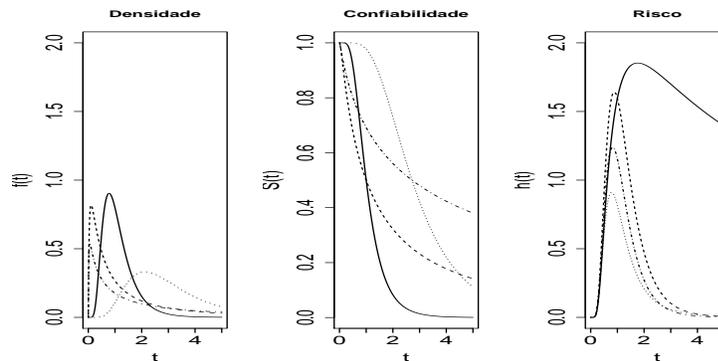


Figura 3: Forma típica das funções de densidade de probabilidade, confiabilidade e risco considerando  $\mu = 0$  e  $\sigma = 0,5$ (—);  $\mu = 0$  e  $\sigma = 1,5$ (--);  $\mu = 1$  e  $\sigma = 0,5$ (...);  $\mu = 1$  e  $\sigma = 2$ (.-.), da distribuição log-normal.

### 3 Inferência Bayesiana

Os procedimentos clássicos da inferência estão construídos com base no princípio da repetibilidade amostral, ou seja, no princípio de que o experimento realizado para obter informações sobre o parâmetro pode ser repetido, sob as mesmas condições, um número arbitrário de vezes. Além disso, um aspecto importante da escola clássica reside em reconhecer a variabilidade que se verifica de amostra para amostra, e isto é quantificado através de uma distribuição amostral  $F_o$ .

Na escola Bayesiana também se quer fazer inferências sobre um parâmetro  $\theta$ , que, da mesma forma, será baseada na informação trazida pelos dados fornecidos pelo experimento realizado. Entretanto, usa-se do experimento apenas a informação trazida pela amostra de fato observada. Ou seja, amostras que poderiam ter sido observadas, mas não foram, não são levadas em conta no processo de inferência.

A incerteza é mensurada através de medidas de probabilidade subjetivas sobre  $\theta$ . Por postura subjetivista entende-se uma forma de pensar na qual assume-se que tem incerteza sobre tudo o que é desconhecido e essa incerteza é quantificada probabilisticamente. Essa informação que se dispõe então, *a priori*, é representada probabilisticamente pela distribuição *a priori* de  $\theta$ , denotada por  $\pi(\theta)$ .

Via teorema de Bayes obtém-se a distribuição *a posteriori* de  $\theta$  dada por:

$$\pi(\theta|\mathbf{y}) = \frac{L(\theta|\mathbf{y})\pi(\theta)}{\int_{\Theta} L(\theta|\mathbf{y})\pi(\theta)d\theta} \quad (1)$$

em que  $\mathbf{y}$  é o vetor de dados observados  $n \times 1$ , e  $L(\theta|y)$  é a função de verossimilhança de  $y$ .

A distribuição *a posteriori* por si só já é uma ótima inferência, pois a partir dela perceberemos as regiões de mais alta densidade, se há multimodalidade, se há assimetria, o peso da cauda, entre outros aspectos. Aliado a isso, os resumos da distribuição *a posteriori* também são úteis e dentre os mais usados se destacam a média, a moda e a mediana *a posteriori*.

Estes resumos são escolhidos em função da forma da distribuição *a posteriori*, ou do quanto se quer perder.

Suponha que  $\theta$  seja um vetor tal que  $\theta = (\theta_1, \dots, \theta_k)$ .

- A média *a posteriori* de  $\theta$  é:

$$\hat{\theta}_{ME} = E(\theta|\mathbf{y}) = (E(\theta_1|\mathbf{y}), \dots, E(\theta_k|\mathbf{y}))$$

- A mediana *a posteriori* de  $\theta$  é:

$$\hat{\theta}_{MD} = Md(\theta|\mathbf{y}) = (Md(\theta_1|\mathbf{y}), \dots, Md(\theta_k|\mathbf{y}))$$

- A moda *a posteriori* de  $\theta$  é o vetor  $\hat{\theta}_{Mo}$ :

$$\pi(\hat{\theta}_{Mo}|\mathbf{y}) = \max_{\theta \in \Theta} \pi(\theta|\mathbf{y}) = \max_{\theta \in \Theta} \pi(\theta)f(\mathbf{y}|\theta)$$

O problema geral da inferência Bayesiana consiste em calcular a distribuição *a posteriori*, devido a integral que aparece no denominador da expressão (3).

O método de Monte Carlo via cadeias de Markov (MCMC) é uma alternativa para se obter a distribuição *a posteriori*. Sua idéia consiste em obter uma amostra da distribuição *a posteriori*, baseada apenas em seu núcleo, e calcular estimativas amostrais de características desta distribuição, podendo ser estabelecido pelo amostrador de Gibbs e o Algoritmo de Metropolis-Hastings. O amostrador de Gibbs será usado durante o trabalho e possui a seguinte formulação:

Suponha que o vetor de interesse é  $\theta = (\theta_1, \dots, \theta_p)$  e  $\mathbf{y} = (y_1, \dots, y_n)$ . Definamos as seguintes distribuições condicionais:

$$[\theta_1|\theta_2, \dots, \theta_p, \mathbf{y}]$$

$$[\theta_2|\theta_1, \theta_3, \dots, \theta_p, \mathbf{y}]$$

...

$$[\theta_p|\theta_2, \dots, \theta_{p-1}, \mathbf{y}]$$

A idéia por trás do amostrador de Gibbs é que podemos criar um algoritmo de simulação da cadeia de Markov da distribuição conjunta *a posteriori*, simulando parâmetros individuais do conjunto de  $p$  distribuições condicionais completas. Simulando um valor de cada parâmetro individual desta distribuição, por sua vez é chamado um ciclo do amostrador de Gibbs. Em condições gerais, este algoritmo de simulação irá convergir para a distribuição alvo (*a posteriori* conjunta de  $\theta$ ) de interesse. (ver Gamerman (1997))

O teste de Geweke (1992) utiliza a comparação dos valores iniciais e finais da cadeia de Markov como um diagnóstico de convergência, com base em um teste para igualdade das médias da primeira e da última parte da cadeia de Markov (em geral, os 10% primeiros e os últimos 50%). Se as amostras são retiradas da distribuição estacionária da cadeia, as duas médias serão iguais e a estatística de Geweke tem uma distribuição normal padrão assintótica.

A estatística  $Z$  de Geweke é a diferença dessas duas médias dividida pelo seu erro padrão assintótico. Como o tamanho da cadeia tende a infinito,  $Z$  pertencerá à uma distribuição normal padrão se a cadeia convergir. Portanto, valores de  $Z$  que caem nos extremos das caudas da distribuição normal padrão sugerem que a cadeia ainda não convergiu.

Outra alternativa popular para o diagnóstico da convergência é o gráfico do traço, que é o gráfico do número da iteração *versus* o valor do parâmetro em cada iteração. Se a cadeia oscilar muito ao longo do espaço paramétrico é um indicador de não convergência.

A Inferência Bayesiana tem sido utilizada na resolução de problemas nas mais diversas áreas do conhecimento. A vantagem em relação à inferência clássica é a possibilidade de incluir na análise não só as informações oriundas da amostra, mas de experimentos passados e experiências de especialistas, através de uma probabilidade subjetiva.

## 4 Modelos de degradação

Neste trabalho utilizamos o método analítico e o método aproximado para modelar os dados de degradação de lasers apresentados, sob a suposição de uma função linear sem intercepto. Essa suposição é plausível, pois na Figura 1 observamos que as degradações de todos os lasers seguem uma função linear e que no tempo zero, essa degradação também é zero.

A degradação para a  $i$ -ésima unidade, no tempo  $t$ , pode ser escrita como:

$$y_{ij} = D_{ij} + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m. \quad (2)$$

$$\epsilon_{ij} \sim Normal(0, \sigma_\epsilon^2)$$

em que  $D_{ij}$  é a degradação da unidade  $i$ , no tempo  $t_j$  e  $\epsilon_{ij}$  é o desvio residual da degradação da unidade  $i$ , no tempo  $t_j$ . A degradação  $D_{ij}$  é função do tempo, e de um parâmetro desconhecido  $\theta$  que diferenciará as  $i$  unidades, podendo ser reescrita como:

$$D_{ij}(t) = D(t_{ij}, \theta) = (1/\theta_i)t_{ij}$$

### 4.1 Método Analítico

Considerando o modelo geral (2), quando a unidade  $i$  atinge o nível crítico de degradação,  $Df$ , a unidade é considerada falha.

$$Df = D_{ij}(t)$$

$$Df = (1/\theta_i)t_{ij}$$

$$T_i = \theta_i * Df$$

Como  $\theta_i$  varia para cada unidade ele é uma variável aleatória e conseqüentemente associa-se uma distribuição de probabilidade a ele. A escolha da distribuição da variável latente  $\theta_i$  determinará a distribuição da variável aleatória  $T_i$ .

Neste trabalho assumimos distribuição Weibull com parâmetros  $\beta$  e  $\lambda$  para o parâmetro  $\theta$ , pois conforme analisado na Seção 5 parece plausível assumir a distribuição Weibull para o tempo até a falha  $T_i$ . Desta forma, o tempo até a falha também possuirá distribuição Weibull com parâmetros  $\beta$  e  $\frac{\lambda}{Df^\beta}$ .

$$\theta_i \sim Weibull(\beta, \lambda)$$

$$F_{T_i}(t) = P(T_i \leq t) = P(\theta_i * Df \leq t) = P\left(\theta_i \leq \frac{t}{Df}\right)$$

$$F_{\theta_i}(t_i/Df) = 1 - \exp \left\{ -\lambda \left( \frac{t}{Df} \right)^\beta \right\}$$

$$f_{T_i}(t) = \frac{d}{dt} F_{\theta_i}(t_i/Df) = \beta \frac{\lambda}{Df^\beta} t^{\beta-1} \exp \left\{ -\frac{\lambda}{Df^\beta} t^\beta \right\}$$

$$T_i \sim Weibull \left( \beta, \frac{\lambda}{Df^\beta} \right)$$

Portanto, o modelo estatístico final para análise de degradação do laser será dado por:

$$y_{ij} | \theta = \theta_i \sim Normal \left( \frac{1}{\theta_i} t_{ij}, \sigma_\epsilon^2 \right),$$

condicional no desconhecimento de  $\theta$ , a degradação observada para a  $i$ -ésima unidade é normalmente distribuída.

Utilizamos uma abordagem Bayesiana para estimar os parâmetros desconhecidos  $(\beta, \lambda, \sigma_\epsilon^2)$  combinando uma informação *a priori* com a informação dos dados, assumindo as seguintes distribuições *a priori*.

$$\beta \sim Gama(a_\beta; b_\beta)$$

$$\lambda \sim Gama(a_\lambda; b_\lambda)$$

$$\tau \sim Gama(a_\tau, a_\tau)$$

em que  $\tau = \frac{1}{\sigma_\epsilon^2}$ .

Utilizamos métodos MCMC para se obter a distribuição *a posteriori* dos parâmetros através do software estatístico *OpenBUGS*. A distribuição *a posteriori* das quantidades de interesse é dada por:

$$\pi(\beta, \lambda, \tau, \theta | \mathbf{y}) \propto f(\mathbf{y} | \theta, \tau) g(\theta | \lambda, \beta) \pi(\lambda, \beta, \tau)$$

em que  $\mathbf{y} = (y_{11}, \dots, y_{ij})$  e  $\theta = (\theta_1, \dots, \theta_i)$ .

As distribuições condicionais completas são dadas por:

$$\pi(\theta | \beta, \lambda, \tau, \mathbf{y}) \propto f(\mathbf{y} | \theta, \tau) g(\theta | \lambda, \beta)$$

$$\pi(\theta | \beta, \lambda, \tau, \mathbf{y}) \propto \tau^n \exp \left\{ (\beta - 1) \sum_{i=1}^n \log \theta_i - \lambda \sum_{i=1}^n \theta_i^\beta - \tau \sum_{j=1}^n \left[ \sum_{i=1}^m \frac{(y_{ij} - t_{ij}/\theta_i)^2}{2} \right] \right\},$$

$$\pi(\tau | \beta, \lambda, \theta, \mathbf{y}) \propto f(\mathbf{y} | \theta, \tau) \pi(\tau)$$

$$\pi(\tau | \beta, \lambda, \theta, \mathbf{y}) \propto \tau^n \exp \left\{ (a_\tau - 1) \log \tau - b_\tau \tau - \tau \sum_{j=1}^n \left[ \sum_{i=1}^m \frac{(y_{ij} - t_{ij}/\theta_i)^2}{2} \right] \right\},$$

$$\pi(\beta|\theta, \lambda, \tau, \mathbf{y}) \propto g(\theta|\lambda, \beta) \pi(\beta)$$

$$\pi(\beta|\theta, \lambda, \tau, \mathbf{y}) \propto \beta^{n+a_\beta-1} \exp \left\{ (\beta - 1) \sum_{i=1}^n \log \theta_i - \lambda \sum_{i=1}^n \theta_i^\beta - b_\beta \beta \right\},$$

e

$$\pi(\lambda|\theta, \beta, \tau, \mathbf{y}) \propto g(\theta|\lambda, \beta) \pi(\lambda)$$

$$\pi(\lambda|\theta, \beta, \tau, \mathbf{y}) \propto \lambda^{n+a_\lambda-1} \exp \left\{ -\lambda \sum_{i=1}^n \theta_i^\beta - b_\lambda \lambda \right\}.$$

## 4.2 Método Aproximado

Assumindo-se o modelo (2) ajusta-se modelos de regressão linear simples para cada uma das  $i$  unidades, a fim de prever qual o tempo em que o nível crítico de degradação  $Df$  será atingido. Esse tempo predito é associado a falha da  $i$ -ésima unidade e chamado de pseudo tempo de falha.

Desta forma, tem-se uma amostra de  $n$  pseudo tempos de falha que é analisada para obter-se os parâmetros  $(\beta, \lambda)$  em abordagens Bayesiana e frequentista.

Considerando uma abordagem Bayesiana combinaremos a informação da amostra da distribuição do tempo até a falha  $\tilde{t}_i$  com as mesmas distribuições *a priori* utilizadas no método analítico, entretando  $\theta_i$  e  $\sigma_\epsilon^2$  não são mais variáveis aleatórias.

Desta forma, a distribuição *a posteriori* das quantidades de interesse é dada por:

$$\pi(\beta, \lambda | \tilde{\mathbf{t}}) \propto f(\tilde{\mathbf{t}} | \lambda, \beta) \pi(\lambda, \beta)$$

em que  $\tilde{\mathbf{t}} = (\tilde{t}_1, \dots, \tilde{t}_i)$  são os pseudo tempos de falha.

As distribuições condicionais completas neste caso são dadas por:

$$\begin{aligned} \pi(\beta | \lambda, \tilde{\mathbf{t}}) &\propto f(\tilde{\mathbf{t}} | \lambda, \beta) \pi(\beta) \\ \pi(\beta | \lambda, \tilde{\mathbf{t}}) &\propto \beta^{n+a_\beta-1} \exp \left\{ (\beta - 1) \sum_{i=1}^n \log \tilde{t}_i - \lambda \sum_{i=1}^n \tilde{t}_i^\beta - b_\beta \beta \right\}, \end{aligned}$$

e

$$\begin{aligned} \pi(\lambda | \beta, \tilde{\mathbf{t}}) &\propto f(\tilde{\mathbf{t}} | \lambda, \beta) \pi(\lambda) \\ \pi(\lambda | \beta, \tilde{\mathbf{t}}) &\propto \lambda^{n+a_\lambda-1} \exp \left\{ -\lambda \sum_{i=1}^n \tilde{t}_i^\beta - b_\lambda \lambda \right\}. \end{aligned}$$

Em uma abordagem frequentista, encontraremos os estimadores de máxima verossimilhança da distribuição Weibull através da amostra de pseudo tempos de falha, a fim de estimar a distribuição do tempo até a falha. (ver Casella (2010))

## 5 Estudo de caso

Apresentamos uma replicação da análise dos dados da degradação de Lasers, apresentados na Seção 4, realizada em Hamada (2005).

Para checar a adequação da suposição da distribuição Weibull para distribuição do tempo até a falha, construímos os gráficos de sobrevivência estimadas por Kaplan-Meier *versus* as sobrevivências estimadas pelos modelos Weibull e log-normal, considerando os pseudo tempos de falha, obtidos através do método aproximado. Na Figura 3 observamos que o modelo log-normal parece não ser adequado, pois apresenta um desvio maior da reta  $x = y$ . Portanto, parece plausível a análise considerando a distribuição Weibull.

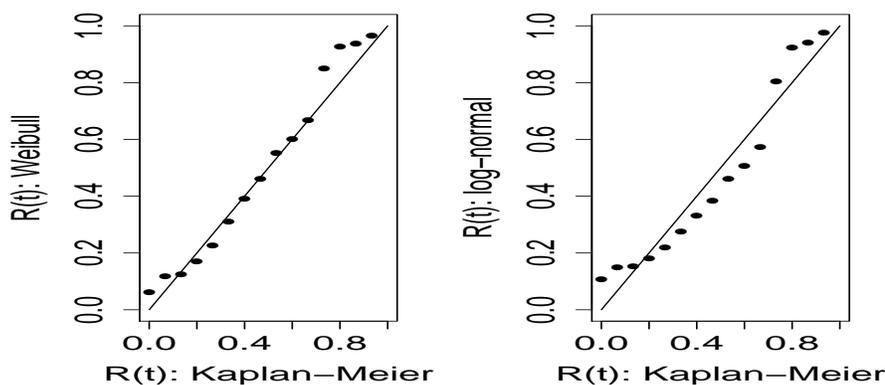


Figura 4: Gráficos das confiabilidades estimadas por Kaplan-Meier *versus* as confiabilidades estimadas pelos modelos Weibull e log-normal, para os pseudo tempos de falha.

Realizamos a análise considerando distribuições *a priori* não informativas para o método analítico e o método aproximado Bayesiano, sendo que os valores dos hiperparâmetros foram dados por  $a_\beta = b_\beta = 0,01$ ,  $a_\lambda = b_\lambda = 0,01$  e  $a_\tau = b_\tau = 100$ .

Na Tabela 1 apresentamos as estatísticas *a posteriori* dos parâmetros  $(\beta, \lambda, \tau)$  para o método analítico e para  $(\beta, \lambda)$  no método aproximado em que a abordagem Bayesiana foi considerada, e as estimativas de máxima verossimilhança para  $(\beta, \lambda)$  no método aproximado considerando a abordagem frequentista. A distribuição *a posteriori* foi obtida através de métodos MCMC, nos softwares estatísticos R e *OpenBUGS*, com período de aquecimento de cadeia igual a 50000, *lag* igual a 50, e tamanho final da amostra igual a 5000.

Algumas funções como a confiabilidade e o percentil  $\alpha$  foram estimadas e comparadas na análise, também são apresentadas na Tabela 1. Para o método analítico e o método aproximado Bayesiano consideramos a média *a posteriori* como estimativas destas funções,

que são dadas a seguir. As variâncias destas funções foram calculadas pelo método delta. (ver Casella (2010))

$$R(t) = \exp \left\{ -(\lambda/Df^\beta)t^\beta \right\}$$

$$t_\alpha = [-(Df^\beta/\lambda) \ln(1 - \alpha)]^{1/\beta}$$

No método aproximado frequentista as estimativas da função de confiabilidade e do percentil  $\alpha$  foram calculadas utilizando a propriedade da invariância do estimador de máxima verossimilhança:

$$\widehat{R}(t) = \exp \left\{ -\widehat{\lambda}t^{\widehat{\beta}} \right\}$$

$$\widehat{t}_\alpha = [-1/\widehat{\lambda} \ln(1 - \alpha)]^{1/\widehat{\beta}}$$

Tabela 1: Estatísticas a posteriori dos dados de degradação de laser.

Método Analítico				
	Média	Desvio Padrão	Intervalo HPD 95%	
			LI	LS
$\beta$	6,5342	1,4238	4,0770	9,5820
$\lambda$	5,01E-11	1,95E-09	4,33E-27	7,84E-12
$\tau$	23,6838	2,2341	19,4700	28,2400
$R(4500)$	0,7524	0,0899	0,5550	0,9019
$t_{0,1}$	3824,00	384,90	3010,00	4510,00
Método Aproximado Bayesiano				
	Média	Desvio Padrão	Intervalo HPD 95%	
			LI	LS
$\beta$	6,0990	1,2170	3,9290	8,792
$\lambda$	7,82E-08	2,15E-06	6,74E-22	1,71E-08
$R(4500)$	0,7347	0,0884	0,5441	0,8833
$t_{0,1}$	3761,00	372,00	2957,00	4413,00
Método Aproximado Frequentista				
	EMV	Desvio Padrão	Intervalo de Confiança 95%	
			LI	LS
$\beta$	6,5990	1,2170	3,9290	8,9320
$\lambda$	2,10E-10	6,67E-48	0,00E+00	4,46E-18
$R(4500)$	0,7622	0,0670	0,6290	0,8529
$t_{0,1}$	3898,6910	259,9493	3471,1120	4326,2690

Observamos na Tabela 1 que o parâmetro de forma da distribuição Weibull,  $\beta$ , para ambos os métodos e abordagens considerados, foi maior que 1. Isso indica que a função risco dos lasers é crescente, ou seja, que as unidades se degradam ao longo do tempo.

Por meio do programa computacional R, foram realizados procedimentos para verificar a convergência das estimativas dos parâmetros de interesse como o diagnóstico de Geweke e o gráfico do traço das estimativas. Na Tabela 2 apresentamos as estatísticas de Geweke para os parâmetros estimados através do método analítico e do método aproximado Bayesiano. Nas Figuras 5 e 6 apresentamos os gráficos para as estatísticas de Geweke e o gráfico do traço e densidade dos parâmetros, respectivamente, para o método analítico. Os mesmos gráficos para o método aproximado Bayesiano são apresentados no Apêndice.

Tabela 2: Estatísticas de Geweke para os parâmetros estimados.

	Método Analítico	Método aproximado
$\beta$	-1,2081	0,7700
$\lambda$	1,5512	-0,0911
$\tau$	-1,3905	-
$R(4500)$	-1,1579	0,6092
$t_{0,1}$	-1,3484	0,5117

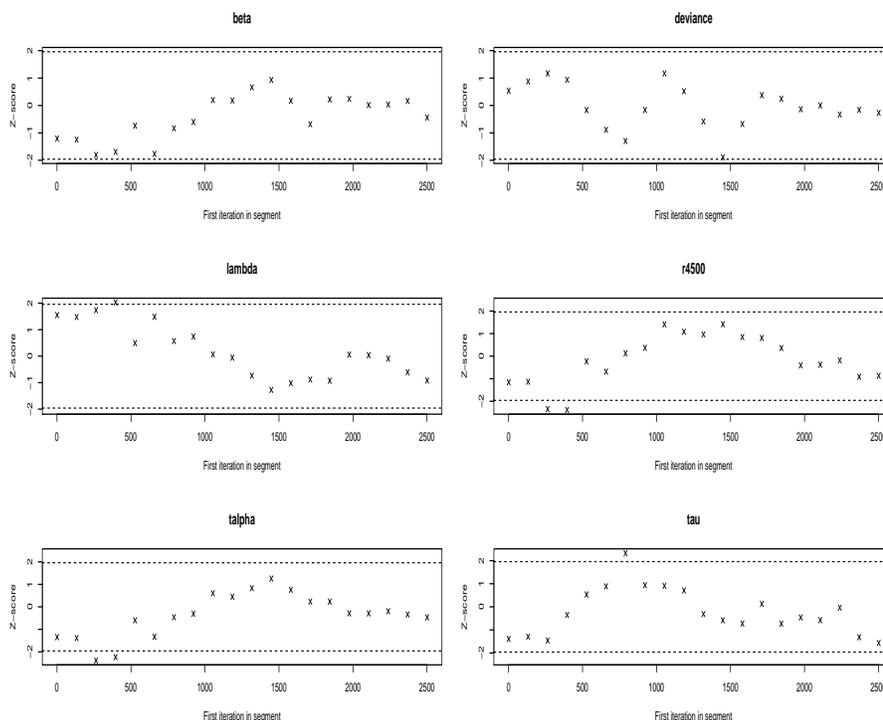


Figura 5: Diagnóstico de convergência para o teste de Geweke para o método analítico.

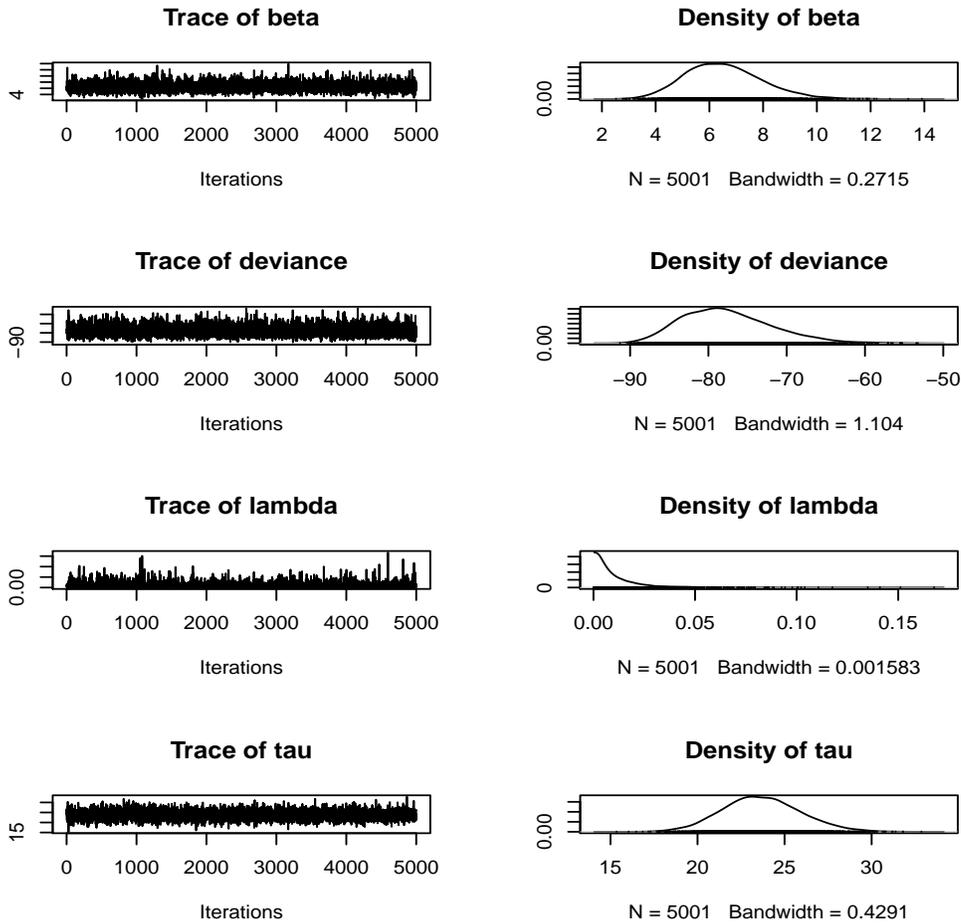


Figura 6: Gráfico do traço e densidade *a posteriori* dos parâmetros  $\beta$ ,  $\lambda$  e  $\tau$  para o método analítico.

A convergência dos parâmetros foi confirmada através das técnicas de diagnósticos utilizadas para ambos os métodos Bayesianos.

Observamos que os dois métodos apresentados no estudo, foram similares nas estimativas dos parâmetros e das funções de interesse, independentemente da abordagem considerada, Bayesiana ou frequentista. Na Figura 7 apresentamos as curvas de confiabilidade para os dois métodos e abordagens consideradas e observamos que suas estimativas foram próximas, o que era esperado, devido a proximidade das estimativas dos parâmetros.

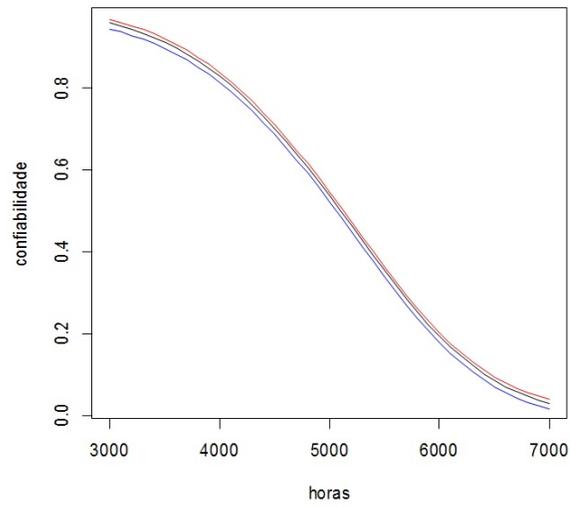


Figura 7: Curvas de confiabilidade dos lasers para o método analítico (preto), o método aproximado Bayesiano (azul) e o método aproximado frequentista (vermelho).

## 6 Considerações Finais

Apresentamos neste trabalho um estudo de dados de degradação mostrando como essa pode ser uma alternativa eficiente em confiabilidade quando se tem pouca informação acerca do tempo de falha. Replicamos o estudo realizado por Hamada 2005 e obtivemos resultados similares. Entretanto, observamos que o autor utilizou um período de aquecimento de cadeia pequeno e não utilizou nenhum *lag* para a geração da distribuição *a posteriori* através de métodos MCMC, o que poderia comprometer os resultados finais.

Com base nos modelos ajustados foram obtidas medidas importantes no estudo de confiabilidade, como a função de confiabilidade e percentis da distribuição do tempo de vida dos produtos. Foi observado que, para o conjunto de dados analisado, os métodos analítico e aproximado apresentaram resultados similares, independente da abordagem, frequentista ou Bayesiana, adotada.

## 7 Apêndice

### 7.1 Gráficos de Diagnóstico

Nas Figuras 8 e 9 são apresentados os gráficos para as estatísticas de Geweke e o gráfico do traço e densidade dos parâmetros, respectivamente, para o método aproximado para os parâmetros  $\beta$  e  $\lambda$ .

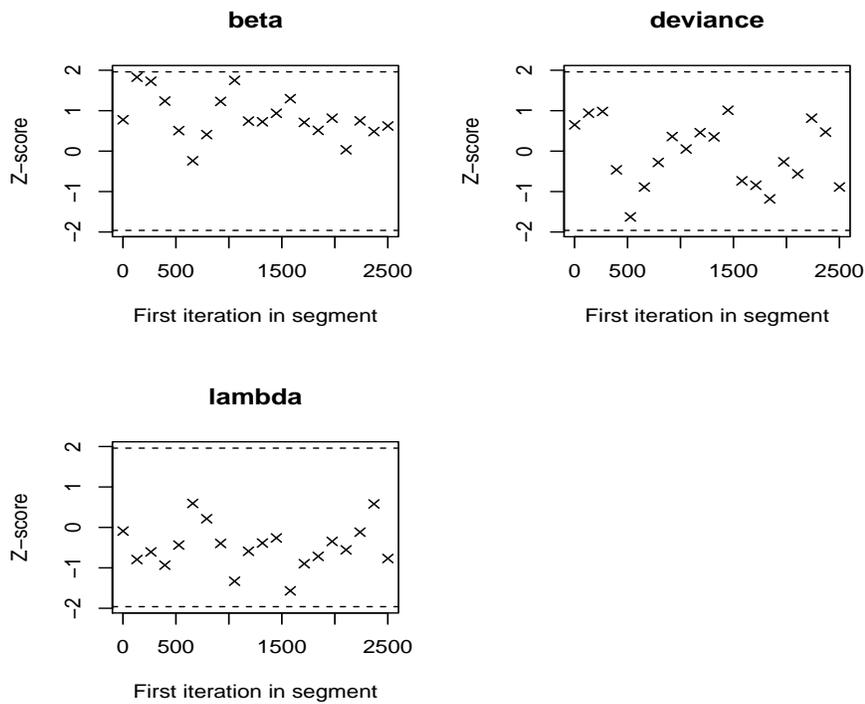


Figura 8: Diagnóstico de convergência para o teste de Geweke para o método aproximado.

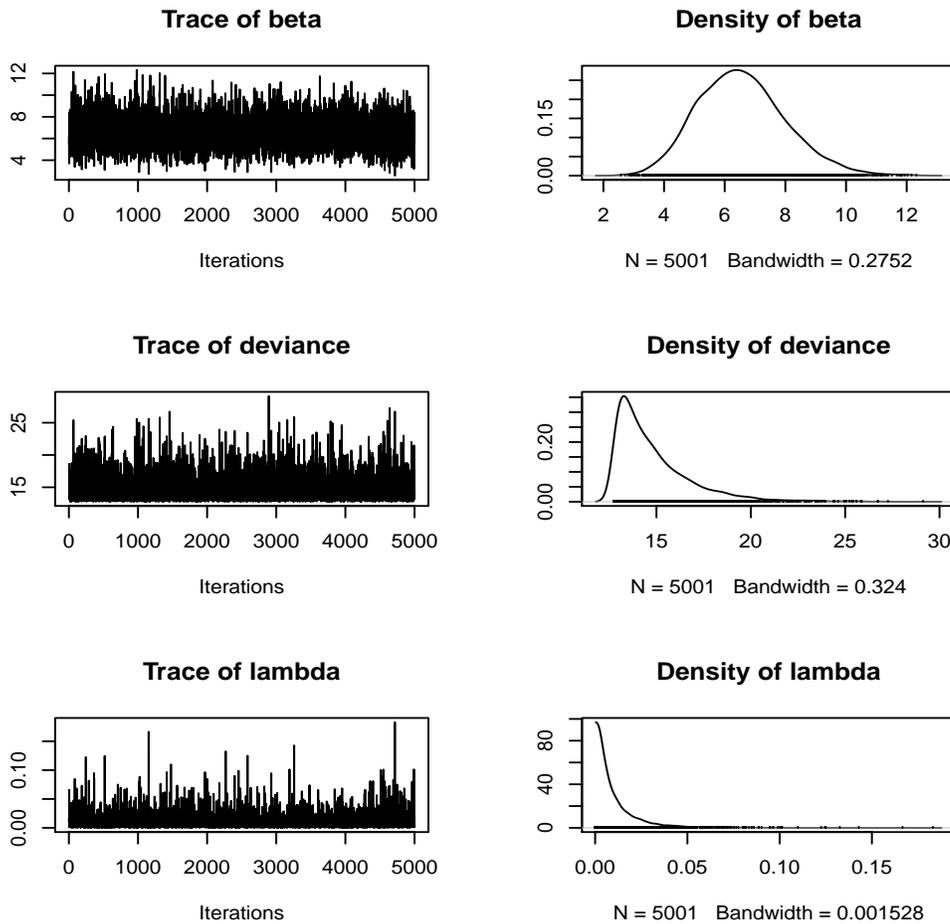


Figura 9: Gráfico do traço e densidade *a posteriori* dos parâmetros  $\beta$  e  $\lambda$  para o método aproximado.

## 7.2 Código em *OpenBUGS*

A seguir é apresentado o código do modelo utilizado para os procedimentos do método analítico Bayesiano, no software *OpenBUGS*.

```

model
{
    for(i in 1:n1)
    {
        y[i]~dnorm(mu[i],tau)
        mu[i]<-hours[i]/igam[units[i]]
    }
}

```

```

}

for( i in 1 : n2 ) {
  igam[i]~dweib(beta,lambda) #(shape,scale)
}

beta~ dgamma(0.01,0.01)
lambda ~ dgamma(0.01,0.01)
tau~dgamma(0.01,0.01)

d<-lambda/pow(L,beta)
r4500<-exp(-d*pow(4500/250,beta))

talpha <- pow(( -(pow(L,beta)/lambda)*log(1-alpha)), (1/beta))

for( i in 1 : npred ) {
  rpred[i]<-exp(-d*pow(tpred[i],beta))
}

}

```

O código abaixo é referente ao modelo utilizado no software *OpenBUGS* para os procedimentos do método aproximado Bayesiano.

```

model

{

for(i in 1:n)
{
t[i] ~ dweib(beta, lambda)
}

beta ~ dgamma(0.001,0.001)
lambda ~ dgamma(0.001,0.001)

d<-lambda/pow(L,beta)
r4500<-exp(-d*pow(4500,beta))

```

```
talpha <- pow(( -(pow(L,beta)/lambda)*log(1-alpha)), (1/beta))

for( i in 1 : npred ) {
  rpred[i]<-exp(-d*pow(tpred[i],beta))

}

}
```

## 8 Referências Bibliográficas

- [1] HAMADA, M. Using Degradation to Assess Reliability. *Quality Engineering*, 17, 615-620, 2005.
- [2] MEEKER, W.Q.; ESCOBAR, L.A. *Statistical Methods for Reliability Data*. Wiley:John Wiley and Sons, Inc, 1998
- [3] HAMADA, M. S.; WILSON, A. G.; REESE, C. S.; MARTZ, H. F. *Bayesian Reliability*. Springer, 2008
- [4] FREITAS, M. A.; COLOSIMO, E. A.; SANTOS, T. R.; PIRES, M. C.: Reliability assessment degradation models: bayesian and classical approaches. *Pesquisa Operacional*, v.30, n.1, p.195-219, Janeiro a Abril de 2010.
- [5] CASELLA, G.; BERGER, R. L. *Inferência Estatística*. São Paulo: Cengage Learning, 2010. 588 p.
- [6] GAMERMAN, D. *Markov Chain Monte Carlo: sthochastic simulation for Bayesian inference*. London; New York, USA: Chapman Hall, 1997. 245 p.
- [7] COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. São Paulo: Blucher, 2006. 367 p.
- [8] GEWEKE, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Oxford: Oxford University Presse*, 169-193, 1992.

## **Efeito de bordadura em experimentos conduzidos no esquema Split Block envolvendo a cultura da soja**

**Guido Gustavo Humada-González<sup>1 3</sup>**

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

**Teodoro Adriano Bruzi<sup>3 3</sup>**

**Gilberto Rodrigues Liska<sup>4 3</sup>**

**Líder Ayala Aguilera<sup>5 3</sup>**

**Jose Rodolfo Humada Sosa<sup>6 3</sup>**

Em uma parcela experimental é frequente a ocorrência de diferenças entre o desempenho de plantas das fileiras laterais e centrais, configurando o efeito de bordadura. Quando esse efeito é distinto para os diversos tratamentos, causa diferença de competição entre parcelas adjacentes. Esse fenômeno tem sido relatado por diversos autores e é um dos diversos fatores que mais afetam a precisão experimental. Objetivou-se avaliar o efeito da bordadura em dez cultivares de soja convencional e transgênica. O experimento foi desenvolvido em faixas em delineamento em blocos casualizados com três repetições, no ciclo agrícola 2013/2014, na fazenda Experimental da Universidade Federal de Lavras – UFLA, à 21° 14' latitude Sul, 45° 00' longitude Oeste em um solo classificado como Latossolo Vermelho-Amarelo. No momento da coleta foram identificadas as linhas da parcela, esse procedimento permitiu que fossem realizadas as análises das duas linhas da área útil e das duas linhas de bordadura. Os fatores estudados foram os cultivares e as linhas, em tanto que a característica agrônômica avaliada foi a produtividade da cultura expressas em kg/ha. De acordo com os resultados obtidos na análise de variância não foram detectadas diferenças significativas, em termos de produtividade, entre as linhas nem entre a interação cultivares com linhas ao nível de significância de 5%. Concluiu-se que não é necessário utilizar bordaduras em experimentos com soja, o que acarreta redução da área experimental.

**Palavras-chave:** *Competição, Experimentos em faixas, Produtividade.*

---

<sup>1</sup> DEX-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [gustavohumad@hotmail.com](mailto:gustavohumad@hotmail.com)

<sup>2</sup> DEX-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [armorais@dex.ufla.com.br](mailto:armorais@dex.ufla.com.br)

<sup>3</sup> DAG-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [adrianobruzi@dag.ufla.br](mailto:adrianobruzi@dag.ufla.br)

<sup>4</sup> DEX-UFLA – Universidade Federal de Lavras, Lavras – MG Email: [gilbertoliska@hotmail.com](mailto:gilbertoliska@hotmail.com)

<sup>5</sup> FCA-UNA – Universidad Nacional de Asunción, San Lorenzo – PY Email: [semillas@agr.una.py](mailto:semillas@agr.una.py)

<sup>6</sup> FCA-UNA – Universidad Nacional de Asunción, San Lorenzo – PY Email: [jochumad@hotmail.com](mailto:jochumad@hotmail.com)

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

# Análise do perfil dos candidatos do ENEM em 2011 por meio de inferência não-paramétrica no município de Uberlândia-MG

Thacyo Euqueres De Villa<sup>1 5</sup>

Douglas Silva Oliveira<sup>2 5</sup>

João Cláudio Quirino Cardoso<sup>3 5</sup>

Janser Moura Pereira<sup>4 5</sup>

O presente estudo tem como objetivo verificar se há diferença significativa no rendimento escolar dos candidatos por meio das notas do Exame Nacional do Ensino Médio (ENEM), em relação à dependência administrativa das escolas e em função do gênero, no município de Uberlândia-MG. Para a pesquisa foram utilizados dados do ENEM, obtidos a partir do portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) referentes ao ano de 2011. Uma vez que a pressuposição de normalidade não foi atendida, aplicou-se o teste não-paramétrico de Kruskal-Wallis, ao nível de significância de 5%, conclui-se que há diferença significativa entre as medianas das notas em relação ao sexo do candidato e a dependência administrativa das escolas.

**Palavras-chave:** *Rendimento escolar, Teste não-paramétrico, Kruskal-Wallis.*

---

<sup>1</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: *thacyo@est.ufu.br*

<sup>2</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: *douglasso1988@gmail.com*

<sup>3</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: *joaoclaudioq@gmail.com*

<sup>4</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: *janser@famat.ufu.br*

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

## Predição do índice combinado de seguradoras utilizando redes neurais artificiais: um estudo introdutório

Richard Hernani Pereira<sup>1 5</sup>

Leandro Ferreira<sup>2 5</sup>

Gustavo Oliveira Marques<sup>3 5</sup>

Marçal Serafim Candido<sup>4 5</sup>

Apresenta-se as seguradoras, sua história no Brasil e a necessidade de acompanhamento por índices. O índice combinado analisa a estrutura dos custos em relação à receita de prêmios. Objetiva-se prever o índice combinado de seguradoras baseando-se nos indicadores econômicos: Índice Geral de Preços do Mercado, Produto Interno Bruto e a Selic, utilizando modelos baseados em redes neurais artificiais. RNA's são estruturas computacionais semelhantes ao sistema nervoso humano e tem capacidade de aquisição e manutenção do conhecimento e gerar resultados. Os dados são anuais de 2003 a 2013. Dividiu-se os dados em grupos; o grupo A para treinamento e teste e o grupo B para predição do índice combinado após as redes treinadas. Empregou-se três tipos de arquiteturas diferentes: (1) 3 neurônios de entrada, 2 na camada escondida e 1 neurônio na saída; (2) 3 neurônios de entrada; 10 na camada escondida e 1 na saída; (3) 3 neurônios na entrada, 20 na camada escondida e 1 na saída, pelo software *FannTool* 1.0. As estimativas são próximas para as arquiteturas. A arquitetura 1 apresentou melhor resultado com valores menores de desvio padrão e variância para o conjunto de dados de treinamento, teste e de predição utilizando o conjunto B.

**Palavras-chave:** *Redes Neurais Artificiais, Seguradoras, Índice Combinado*

<sup>1</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [richard.hernani@gmail.com](mailto:richard.hernani@gmail.com)

<sup>2</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [leandro.ferreira@unifal-mg.edu.br](mailto:leandro.ferreira@unifal-mg.edu.br)

<sup>3</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [gustavo-omarques@outlook.com](mailto:gustavo-omarques@outlook.com)

<sup>4</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [marcal.unifal@yahoo.com.br](mailto:marcal.unifal@yahoo.com.br)

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

# PERFIL ESTATÍSTICO SOBRE SAÚDE E QUALIDADE DE VIDA DA COMUNIDADE ESCOLAR DO IFNMG *CAMPUS* MONTES CLAROS: IMPLEMENTAÇÃO COMPUTACIONAL PARA ACOMPANHAMENTO EM TEMPO REAL

Daniel Antunes Rocha <sup>1</sup>

Bárbara de Oliveira Aguiar<sup>2</sup>

Laércio Ives Santos<sup>3</sup>

Maria de Fátima Ferreira Almeida<sup>4</sup>

Uma das características dos seres humanos é a necessidade de viver bem e sempre buscar condições que melhorem o cotidiano e que permitam uma qualidade de vida satisfatória. Qualidade de vida envolve várias áreas, entre elas o bem físico, mental, psicológico e emocional... Ela se constitui em um dos objetivos a ser alcançado no presente estágio de desenvolvimento da humanidade [M. R. C. Nobre]. O projeto PROVIDA SAUDAVEL pretende coletar e analisar dados provenientes de avaliações físicas por meio do questionário IPAC apresentado em formato web, com o objetivo de disponibilizar o perfil físico e hábitos da população acadêmica em tempo real. Entre as linguagens utilizadas estão: HTML, linguagem para descrever a estrutura das páginas web [w3c.org], Java Script, PHP, CSS, e ainda utilizará alguns frameworks. A implementação se organiza por criação de layout, implementação de pré-processamento de hipertexto e criação de banco de dados, buscando intervir positivamente na sociedade e buscando o melhor uso da web. O projeto possibilitará a intervenção de nutricionista, educador físico e médico do IFNMG na intenção de auxiliar na busca por uma melhor qualidade de vida e inferir na melhoria dos hábitos utilizando recursos advindos da estatística descritiva. O uso da estatística trará relatórios aos usuários permitindo comparações com a estatística em geral, viabilizando políticas de intervenção no campus por parte dos profissionais de saúde.

**Palavras-chave:** *Qualidade de Vida, Pagina Web, Saúde.*

---

<sup>1</sup>IFNMG: Instituto Federal do Norte de Minas Gerais. Email: *danielantrocha@gmail.com*

<sup>2</sup>IFNMG: Instituto Federal do Norte de Minas Gerais. Email: *barbara.oliveira.aguiar@gmail.com*

<sup>3</sup>IFNMG: Instituto Federal do Norte de Minas Gerais. Email: *laercio.santos@ifnmg.edu.br*

<sup>4</sup>IFNMG: Instituto Federal do Norte de Minas Gerais. Email: *fatima.almeida@ifnmg.edu.br*

# **Determinantes do indicador cobertura das provisões matemáticas dos fundos de pensão: um estudo econométrico no período de 2010 a 2013.**

**Luis Henrique Moreira Gonçalves Araújo<sup>1 3</sup>**

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

## **RESUMO**

Os fundos de pensão são formados pelo regime de capitalização que considera na fixação do compromisso futuro de pagamentos de benefícios, a base para o estabelecimento e fixação das contribuições. Neste contexto, sua análise é importante, pois, possibilitam maior controle sobre os riscos a que os fundos estão sujeitos e maximizam a eficiência na tomada de decisões. Diante disso, o objetivo do trabalho foi identificar os determinantes da cobertura das provisões matemáticas. Para isso, foram utilizados modelos de correlação de Pearson e regressão linear múltipla. Foram selecionados os 100 maiores fundos de pensão do Brasil em quantidade de investimentos entre os anos de 2009 e 2012. A partir do ajuste do modelo de regressão, constatou-se que o conjunto de variáveis independentes conseguem explicar as variações na variável dependente em, aproximadamente, 78% dos casos, com um nível de significância inferior a 1%. A partir do teste t, constatou-se que todas as variáveis independentes foram significativas. Quanto aos coeficientes, a variável taxa de administração efetiva apresentou o valor mais expressivo, aproximadamente 0,31, o que denota uma relação direta e menos do que proporcional. A variável taxa de administração apresentou uma relação inversa, porém menos que proporcional (-0,18). Em síntese, os relacionamento entre as variáveis estão coerentes com a teoria de seguros e os resultados apresentados neste trabalho podem auxiliar gestores no processo de tomada de decisão.

**Palavras-chave:** seguros, regressão, fundos de pensão.

---

<sup>1</sup>UNIFAL: Universidade Federal de Alfenas/campus Varginha, Varginha-MG. Email: luishenriquemga@hotmail.com.

<sup>2</sup>UNIFAL: Universidade Federal de Alfenas/campus Varginha, Varginha-MG. Email: gabriel.pessanha@unifal-mg.edu.br.

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

# Estudo de Confiabilidade de Motores Diesel de Caminhões Fora de Estrada

Adriano Gonçalves dos Santos Ribeiro<sup>1</sup>

Gean Carlo Feliciano de Almeida<sup>2</sup>

Graziela Dutra Rocha Gouvêa<sup>3</sup>

Este trabalho consiste em ajustar um modelo de regressão paramétrico que melhor explique o tempo de vida útil de motores de caminhões fora de estrada de uma grande empresa de minério de ferro. No estudo são utilizados dados de 64 motores de um complexo minerador composto por 4 minas (os tempos de vida útil dos motores possuem dados censurados). Através do teste de razão de verossimilhança utilizando a distribuição gama generalizada verificou-se entre as distribuições Weibull, Exponencial e Log-normal, a que melhor se ajusta aos tempos de vida útil dos motores. Os resultados mostraram que a distribuição Weibull melhor se ajustou aos dados, desta forma, foi realizado o ajuste de um modelo de regressão paramétrico Weibull, levando em consideração as variáveis: mina em que os motores tiveram seu regime de operação, inspetor responsável, tipo de manutenção (preventiva ou corretiva) e local de reforma do motor. Para o modelo proposto as variáveis “mina” na presença de “local de reforma” mostraram-se significativas para o modelo. A análise dos resíduos mostrou não existir evidências de violação para o modelo proposto. Os resultados encontrados forneceram estimativas importantes para os gestores e analistas de frota da empresa

**Palavras-chave:** confiabilidade de motores, caminhões fora de estrada.

---

<sup>1</sup>ICT – UFOP: Universidade Federal de Ouro Preto-MG. Email:gsribeiro2003@yahoo.com.br

<sup>2</sup>ICT – UFOP: Universidade Federal de Ouro Preto-MG. Email:geanfeliciano@yahoo.com.br

<sup>3</sup>ICT – UFOP: Universidade Federal de Ouro Preto-MG. Email: gragouvea@gmail.com

## Estabilidade fenotípica de genótipos de soja tolerantes ao oídio

Éder Matsuo<sup>1 2</sup>

Ronaldo Machado Júnior<sup>1 2</sup>

A utilização de cultivares de soja tolerantes ao oídio é uma das medidas de controle de maior eficiência. Objetivou-se identificar genótipos de soja tolerantes ao oídio com auxílio de procedimentos de estabilidade fenotípica. Em casa de vegetação, foram analisados 25 genótipos por meio da inoculação artificial do patógeno e avaliação da severidade em nove semanas consecutivas, a partir dos primeiros sintomas da doença. Adotou-se o delineamento inteiramente casualizado, com cinco repetições, no esquema parcela subdividida (parcelas representadas pelos genótipos e subparcelas pelas avaliações). Os dados foram analisados quanto à normalidade e, posteriormente, submetidos à análise de variância, agrupamento de Scott e Knott ( $\alpha=0,05$ ) e aos procedimentos Plaisted e Peterson (1959) e Wricke (1965). Identificou-se diferença significativa a 1% pelo teste F para o efeito de parcela, subparcela e interação. O genótipo A5STT-MT comportou-se como tolerante ao patógeno e de alta estabilidade fenotípica, quando comparado com o padrão de tolerância (MG/BR-46 (Conquista)). Assim, foi possível identificar genótipos de soja tolerantes ao oídio.

**Palavras-chave:** *Glycine max*, Melhoramento, Biometria.

---

<sup>1</sup> UFV/CRP: Universidade Federal de Viçosa – Campus de Rio Paranaíba, Rio Paranaíba, MG. E-mails: [edermatsuo@ufv.br](mailto:edermatsuo@ufv.br); [ronaldojr\\_agro@hotmail.com](mailto:ronaldojr_agro@hotmail.com)

<sup>2</sup> Agradecimento à FAPEMIG e ao CNPq pelo apoio financeiro.

# Uma justificativa geométrica para o estimador de James-Stein centrado em um ponto qualquer

Cristiane Alvarenga Gajo<sup>1 3</sup>

Lucas Monteiro Chaves<sup>2 3</sup>

James e Stein (1961), apresentaram o resultado surpreendente de um estimador para a média da normal multivariada, de dimensão maior ou igual que três, com erro quadrático médio uniformemente menor que o estimador de máxima verossimilhança, a média amostral, isto é demonstrar a não admissibilidade da média amostral. Este novo estimador referido na literatura como o estimador de James-Stein, quando centrado na origem, é um caso particular de estimador de encolhimento. Uma justificativa heurística baseada em argumentos geométricos do porque o estimador de James-Stein centrado na origem domina o estimador média é apresentado no artigo original de James e Stein (1961) e de forma bastante detalhada por Brandwein e Strawderman (1990). Esta justificativa basea-se essencialmente em uma construção tipo encolhimento. No entanto, para o estimador de James-Stein centrado em um ponto qualquer o argumento do encolhimento não pode ser aplicado em geral. Neste trabalho uma nova justificativa para o estimador de James-Stein centrado em um ponto qualquer é apresentada.

**Palavras-chave:** *Estimador de James-Stein, Argumentos Geométricos.*

---

<sup>1</sup>UFLA: Universidade Federal Lavras, Lavras-MG. Email: [cristianegajo@yahoo.com.br](mailto:cristianegajo@yahoo.com.br)

<sup>2</sup>UFLA: Universidade Federal Lavras, Lavras-MG. Email: [lucas@dex.ufla.br](mailto:lucas@dex.ufla.br)

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

# Análise Longitudinal do Impacto do Programa Bolsa Família

Aline Araújo Barbosa de Castro<sup>1</sup>

Carolina Ribeiro Antonio<sup>2</sup>

Cínara de Jesus Santos<sup>3</sup>

Tamires Mara Coelho<sup>4</sup>

Walmir dos Reis Miranda Filho<sup>5</sup>

Marcel Toledo de Vieira<sup>6</sup>

Ricardo da Silva Freguglia<sup>7</sup>

O Bolsa Família, criado em 2003, é um programa de transferência direta de renda com condicionantes nas áreas de educação e saúde, que beneficia famílias pobres e extremamente pobres. A Pesquisa de Avaliação de Impacto do Bolsa Família (AIBF) é um estudo longitudinal por amostragem complexa que contrasta famílias beneficiárias e não beneficiárias do programa, com o objetivo de entender qual o seu impacto nas condições de vida das mesmas. O levantamento dos dados foi realizado em 2005 e 2009 e foram definidos 03 grupos de observação – beneficiários, possíveis beneficiários, e não cadastrados em programas sociais. O objetivo deste trabalho é, primeiramente, relatar as dificuldades enfrentadas na preparação e posterior análise de uma base de dados real de uma pesquisa domiciliar longitudinal de grande porte. Os dados da referida pesquisa estão originalmente divididos em arquivos de tamanho considerável onde se observaram dificuldades como confiabilidade por haver dados faltantes e informações conflitantes, por exemplo. Dado o número de variáveis envolvidas houve certa dificuldade em identificar informações pertinentes à análise, definir parâmetros de interesse e também dificuldade quanto aos domínios de análise do estudo. Contornados estes desafios, foram estudadas de forma exploratória as características de consumo referentes à alimentação, saúde, bens duráveis, vestuário e moradia dos grupos de observação, procurando avaliar possíveis impactos do programa bolsa família.

**Palavras-chave:** *Bolsa Família, análise longitudinal, grandes bases de dados.*

---

<sup>1</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. e-mail: alineabcastro@hotmail.com

<sup>2</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. e-mail: carolinaribeiroantonio@hotmail.com

<sup>3</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. e-mail: cinara.uffj@gmail.com

<sup>4</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. e-mail: tamires\_mcoelho@hotmail.com

<sup>5</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. e-mail: walmir.reis.miranda.filho@gmail.com

<sup>6</sup> ICE – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. e-mail: marcel.vieira@ice.uffj.br

<sup>7</sup> Fac. De Economia – UFJF: Universidade Federal de Juiz de Fora, Juiz de Fora-MG. e-mail: ricardo.freguglia@uffj.edu.br

## Breve Perfil da Saúde dos Discentes no Campus da UFMG

Felipe Tiago Pacheco de Freitas<sup>1,2</sup>

Isabela de Lima Severino<sup>1,2</sup>

Janete Furtado Ribeiro Aguirre<sup>1,2</sup>

Zaba Valtuille Lieber<sup>1,2</sup>

De acordo com a Organização Mundial de Saúde, “uma dieta adequada, bem equilibrada combinada com atividade física regular são os pilares da boa saúde”. No presente trabalho, elaborou-se um questionário de pesquisa para identificar de forma superficial, alunos portadores desse padrão de “boa saúde” em 4 unidades do campus UFMG-Pampulha: Instituto de Ciências Exatas, Instituto de Ciências Biológicas, Escola de Engenharia e Faculdade de Ciências Humanas. A coleta de dados foi feita por meio de questionário auto-explicativo, respondido de forma voluntária, sem identificação por parte dos 120 entrevistados. Os dados coletados foram cadastrados em planilha MSEXCEL® e as estatísticas descritivas produzidas no software Minitab17®. As respostas dos entrevistados receberam pontos negativos ou positivos, segundo critérios de classificação de frequência de consumo de certos tipos de alimentos, prática de esportes e prevenção de doenças. Descobriu-se daí que: alunos de engenharia são os que mais praticam atividades físicas; os de ciências humanas e biológicas se assemelham na preocupação com a boa alimentação, enquanto os de ciências humanas são mais conscientes quanto ao cuidado preventivo de doenças. Em geral, os de ciências exatas se posicionaram com valores inferiores aos demais, sendo os que menos se preocupam com a prevenção de doenças.

Os resultados aqui não têm valor inferencial para a população amostrada, visto que o processo de amostragem foi por conveniência.

**Palavras-chave:** *Boa saúde, nutrição, atividades físicas, alunos.*

---

<sup>1</sup> UFMG — Curso de Graduação em Estatística. E-mails: [felipetiago@hotmail.com](mailto:felipetiago@hotmail.com); [ils\\_isabelalima@yahoo.com.br](mailto:ils_isabelalima@yahoo.com.br); [janete.agurre@gmail.com](mailto:janete.agurre@gmail.com); [zaba.lieber@hotmail.com](mailto:zaba.lieber@hotmail.com).

<sup>2</sup> Agradecimento à Profa. Ilka Afonso Reis pela leitura e críticas ao trabalho.

# Aplicação de séries temporais ao PIB Indústria.

**Luiz Renan Toffanin da Silva<sup>1 3</sup>**

**Luciene Resende Gonçalves<sup>2 3</sup>**

## 1- Introdução

Com a complexidade que os mercados têm atingido atualmente, e o crescente interesse pelos dados econômicos dos países, a metodologia de séries temporais tem ajudado na sistematização dos dados e sua abordagem – desde correção de erros a tabulação e geração de perspectivas para os mesmo.

Dessa maneira, o presente trabalho utilizará séries temporais no conjunto de dados referente ao PIB Industrial do Brasil, isolando o setor de Produção e Distribuição de Eletricidade, Gás e Água, insumos de necessidade básica para o crescimento e desenvolvimento de um país.

## 2- Objetivos

Identificar um modelo de previsão para o PIB Indústria – Produção e Distribuição – Eletricidade, Gás e Água do Brasil, por meio da análise de séries temporais, prospectando, assim, cenários futuros para o crescimento e desenvolvimento do país.

## 3- Metodologia

Para a manipulação dos procedimentos estatísticos do presente estudo, foi utilizado o software computacional Gretl, que possibilitou investigar o que estava acontecendo com a série estudada. Para começar a utilização da série seus correlogramas e gráficos foram analisados e apresentaram a

<sup>1</sup> ICSA – UNIFAL/MG – Universidade Federal de Alfenas, campus Varginha, Varginha-MG. Email: [renantoffanin@hotmail.com](mailto:renantoffanin@hotmail.com).

<sup>2</sup> ICSA – UNIFAL/MG – Universidade Federal de Alfenas, campus Varginha, Varginha-MG. Email: [luciene.goncalves@unifal-mg.edu.br](mailto:luciene.goncalves@unifal-mg.edu.br).

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

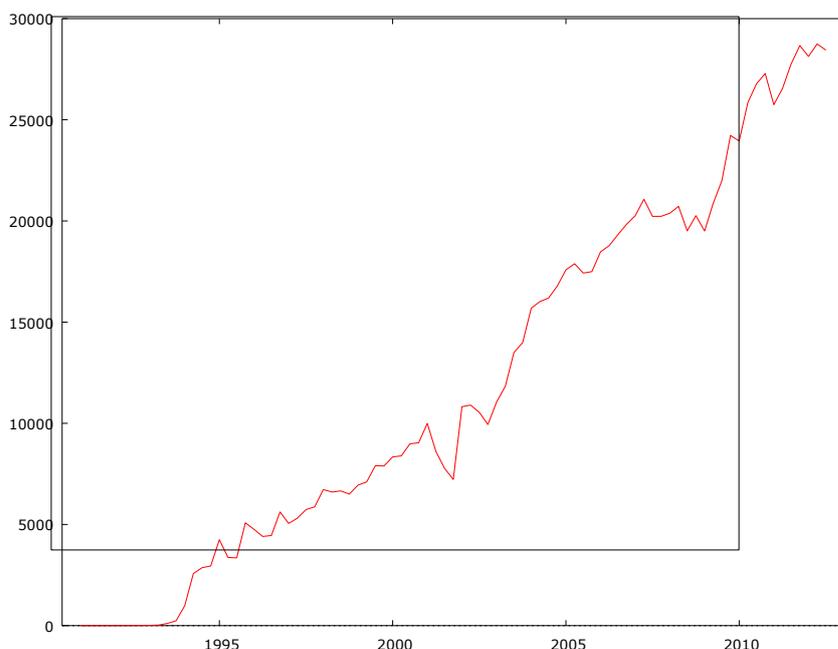
existência de tendência na série. Para corrigir esse problema foi necessário acrescentar a Primeira Diferença Da Variável PIB, a fim de eliminar a sua tendência.

Para comprovar a ocorrência de tendência na série foram realizados os testes de Ljung-Box, Cox-Stuart, e Dicky Fuller aumentado, e também para confirmar que a série não apresentava sazonalidade foi realizado o teste de Fisher, os quais serão apresentados na sessão resultados mostrando seus respectivos gráficos e contas.

Em seguida foi realizado o teste de Box e Jenkins a fim de encontrar as previsões futuras dessa série. De acordo com Box e Jenkins (1976), os dados por si transmitem as informações necessárias para investigar relações e estruturas comportamentais do conjunto de observações da variável proposta, possibilitando gerar estacionariedade na série analisada, o PIB Indústria, Produção e Distribuição - Eletricidade, Gás e Água referente aos anos de 1991 até 2013 analisados trimestralmente, do Brasil.

#### 4- Resultados e discussão

A série original do PIB Indústria- Produção e Distribuição – eletricidade, gás e água, com frequência trimestral de 1991:1 à 2013:3, é dada pelo seguinte gráfico de série temporal, e pelo seu correlograma, respectivamente. Para melhor manejo dos dados, foi definido um intervalo de tempo, referente à 1991:1 à 2012:3.



A partir de sua verificação, foi notado visualmente um componente da serie temporal que se caracteriza pela tendência. A tendência em si impede que a série seja estacionaria e ruído branco.

Com a finalidade de obter previsões para a série, o próximo passo realizado foi a modelagem ARIMA do conjunto de dados. Foram realizados algumas tentativas que não se adequaram aos processos de estacionariedade. Entre os modelos, muitos modelos não apresentavam significância estatística, enquanto outros até apresentavam, porém não eram um ruído branco. Segue a baixo a lista dos modelos testados que não se adequaram :

Modelo ARIMA (p,d,q):

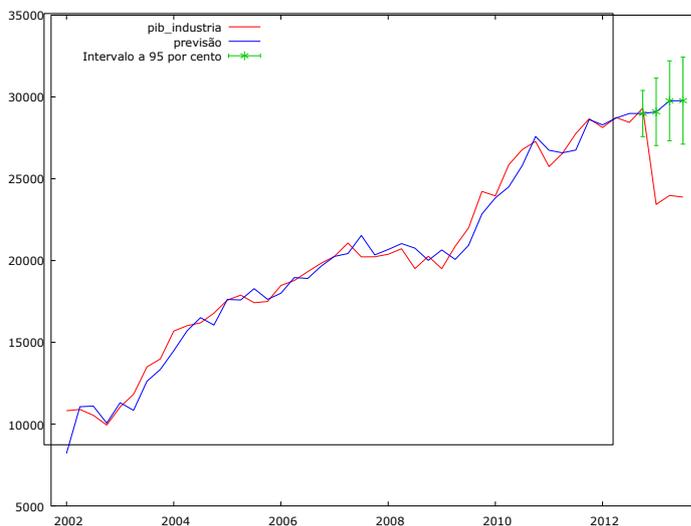
(1,1,1) ; (1,1,2) ; (1,1,3) ;(1,1,4)

(2,1,1) ;(2,1,2);(2,1,3);(2,1,4)

(3,1,1);(3,1,3);(3,1,4)

(4,1,1);(4,1,2);(4,1,3)(4,1,4) (4,1,0) (0,1,4)

Esse primeiro passo pode ser classificado como identificação da série para encontrar os valores adequados. A partir disso foi encontrado o modelo ARIMA (3,1,2), porém foi necessário realizar alguns testes para comprovar sua adequacidade.



A partir dessas previsões foi possível estimar a equação do modelo, como pode ser vista abaixo:

$$\alpha (1 - \alpha 1 B - \alpha 2 B^2 - \alpha 3 B^3) \cdot (1 - B)^1 y_t = \theta + (1 - \beta 1 B) + \beta 2 B^2, \text{ ou seja,}$$

$$(1+1,07510B + 0,460 B^2 + 0,2961 B^3) (1-B) y_t = 333,35 + (1,13823B) + (0,393532B^2) \\ (0,257531) \quad (0,2329) \quad (0,1204) \quad (69,64) \quad (0,2647) \quad (0,2310)$$

## 5- Considerações finais

O presente trabalho teve como objetivo descrever o ajustamento de um modelo de séries temporais. Obtendo os dados a partir de pesquisa do site IPEA- DATA notou-se em primeira instância que os dados coletados apresentavam um componente tendencial, ao qual foram manipulados testes e modelagens afim de tornar essa série estacionária.

Com os testes e aplicações de medias de correção foi possível eliminar a tendência da série, tornando-a um ruído branco. Desse modo, após a serie estar devidamente enquadrada nos pressupostos de séries temporais, foi possível realizar previsões, e estimar seu modelo.

Neste trabalho, foi apresentada uma maneira sumária apenas dos modelos estatísticos mais utilizados em análise de séries temporais, com ajuda de um software adequado foi possível a realização dos testes e modelos.

Desse modo notou-se que o componente estudado PIB Industrial apresenta momentos de intervenção, onde pode ser explicado pelas instabilidades econômicas, e pela sua variação de dados de acordo com o tempo.

## Referencial bibliográfico

Agência Indusnet Fiesp. (06 de 09 de 2013). Participação da indústria no PIB pode cair para 9,3% em 2029, aponta estudo da Fiesp. São Paulo, SP, Brasil.

ANDRADE, João Souza. Apontamentos de Econometria Aplicada. 2001 – (Maio 2004). Disponível em: <http://www4.fe.uc.pt/jasa/estudos/econometria.pdf>. Acesso em: janeiro de 2014

BEZERRA, Manoel Ivanildo Silvestre. Analise de Séries Temporais. São Paulo. 2006. Disponível em: <http://pt.scribd.com/doc/5515941/88/O-Teste-de-Box-Pierce>. Acesso em: janeiro de 2014.

CONCEIÇÃO, A. (28 de 10 de 2013). Analistas baixam previsão para PIB e produção industrial em 2014. São Paulo, SP, Brasil.

ESTADÃO. (28 de 02 de 2002). PIB DE 2001 CRESCE 1,51% SOBRE O ANO DE 200. SÃO PAULO, SP, BRASIL.

GUIMARÃES, Alam Gonçalves e Sáfadi, Thelma. Uso de Séries Temporais na Análise do Número de Acidentes na BR-101 no Trecho do Estado de Santa Catarina. Universidade Federal de Lavras, Minas Gerais. Disponível em:

[http://www.cbtu.gov.br/monografia/2009/trabalhos/artigos/planejamento/5\\_219\\_AC.pdf](http://www.cbtu.gov.br/monografia/2009/trabalhos/artigos/planejamento/5_219_AC.pdf). Acesso em: janeiro de 2014.

IBGE. (28 de 02 de 2002). MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO. Acesso em 2014 de 01 de 28, disponível em INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA: <http://www1.ibge.gov.br/home/presidencia/noticias/28022002pib.shtm>

IMB. (20 de 05 de 2008). PRODUTO INTERNO BRUTO DO ESTADO DE GOIÁS - 2007. Goiás, GO, Brasil.

FRANCO, Glaura C. Apostila sobre Modelagem Box and Jenkins. Universidade Federal de Minas Gerais, Minas Gerais. Disponível em: <ftp://est.ufmg.br/pub/glaura/Met%20Previsao/Modelagem%20Box%20and%20Jenkins.pdf>. Acesso em: janeiro de 2014. (IBGE, 2002)

PEREIRA, Marcelo Bourguignon. Teste de Dickey-Fuller Robusto Baseado nos Ranks para Séries Temporais com Observações Atípicas. Vitória, Espírito Santo. 2010. Disponível em: [http://www.ime.unicamp.br/sinape/sites/default/files/ConcursoIC\\_Monografia.pdf](http://www.ime.unicamp.br/sinape/sites/default/files/ConcursoIC_Monografia.pdf). Acesso em: janeiro de 2014.

PORTAL ACTION. Teste de Box-Pierce. Disponível em: <http://www.portalaction.com.br/1480-471-teste-de-box-pierce>. Acesso em: janeiro de 2014.

NASCIMENTO, M. et al. Teste dos Sinais para Tendência: Uma Aplicação em Melhoramento de Plantas. São Paulo, v.26, n.4, p. 19-30, 2008. Disponível em: [http://jaguar.fcav.unesp.br/RME/fasciculos/v26/v26\\_n4/A2\\_Artigo\\_Moyses.pdf](http://jaguar.fcav.unesp.br/RME/fasciculos/v26/v26_n4/A2_Artigo_Moyses.pdf). Acesso em janeiro de 2014.

OLIVEIRA, S. de. Um estudo em séries temporais na análise da receita nominal de vendas de veículos e motos; Conexão ci.: r. cient. UNIFOR-MG, Formiga, v. 7, n. 2, p. 64-73, jul./dez. 2012

## ESTIMATIVA DO TAMANHO

### ÓTIMO DE PARCELA EM PESQUISA COM TARO.

**Willerson Silva<sup>1</sup>**

**Mário Puiatti<sup>2</sup>**

**Paulo R Cecon<sup>3</sup>**

**Tocio Sedyama<sup>4</sup>**

**Antônio P S Carneiro<sup>5</sup>**

O taro é uma hortaliça tuberosa que produz rizomas de elevado valor nutritivo e energético, com boa conservação pós-colheita. Na pesquisa científica, o tamanho da parcela experimental influencia diretamente a precisão experimental. Para resolver o problema da determinação do tamanho ótimo de parcelas, vários métodos têm sido relatados na literatura. O Método da Máxima Curvatura Modificado (MMCM) tem sido o mais utilizado, porém, métodos fundamentados no ajuste de modelos de regressão segmentada vêm sendo utilizados com sucesso. Por essas razões, objetivou-se utilizar o método Modelo Linear de Resposta Platô (MLRP), e o MMCM, na determinação do tamanho ótimo de parcela para pesquisa com taro. O experimento foi conduzido na Horta do Departamento de Fitotecnia da UFV, no período de set/2011 a jun./2012. Foi considerado como área útil as 20 fileiras centrais com 18 plantas cada, no espaçamento de 1,0 x 0,3 m, totalizando 360 plantas. Feito os possíveis agrupamentos, formaram-se 23 diferentes tamanhos de parcelas. Avaliaram-se as características peso do rizoma filho grande (PRFG) e peso dos rizomas filhos comerciáveis (PRFC). Os métodos MLRP e MCM produziram resultados semelhantes, o tamanho ótimo de parcela experimental indicada foi de 15 plantas o que equivale à área de 4,5 m<sup>2</sup>.

**Palavras-chave:** Tamanho ótimo, Parcela, Taro.

---

<sup>1</sup> IFMG-Instituto Federal de Minas Gerais, Governador Valadares-MG. E-mail: willerson.silva@ifmg.edu.br

<sup>2</sup> UFV-Universidade Federal de Viçosa, Viçosa-MG. E-mail: mpuiatti@ufv.br

<sup>3</sup> UFV-Universidade Federal de Viçosa, Viçosa-MG. E-mail: cecon@ufv.br

<sup>4</sup> UFV-Universidade Federal de Viçosa, Viçosa-MG. E-mail: t.sedyama@ufv.br

<sup>5</sup> UFV-Universidade Federal de Viçosa, Viçosa-MG. E-mail: policarpo.carneiro@gmail.com

# Modelagem de secas ocorridas no estado do Paraná

Raissa Moreira Simão<sup>1 4</sup>

Ana Paula Coelho Madeira Silva<sup>2 4</sup>

Jailson de Araujo Rodrigues<sup>3</sup>

A insuficiência de precipitação pluviométrica ocasiona o fenômeno climático conhecido como seca. A seca está diretamente relacionada aos mais diversos setores da sociedade, afetando a economia, o meio ambiente e a qualidade de vida dos cidadãos. O Índice Padronizado de Precipitação – SPI quantifica o déficit ou o excesso de precipitação em diferentes escalas de tempo e mensura a intensidade da seca ocorrida em uma determinada região. Ajustar diferentes modelos probabilísticos a este índice de quantificação hídrica, verificar e comprovar a sua aderência se mostra uma forma interessante de analisar os valores por ele fornecidos e prover informações de utilidade pública. O objetivo deste trabalho é verificar a aderência dos modelos gama e log normal à dados de seca do estado do Paraná. A estimação dos parâmetros dos modelos foi feita através do método da Máxima Verossimilhança. A aderência foi comprovada analiticamente pelo teste de Komolgorov-Smirnov e graficamente pelo q-qplot. O Critério de Informação de Akaike (AIC) foi utilizado para analisar o modelo que melhor se ajustou aos dados. Os resultados obtidos indicaram que a distribuição gama apresentou ajuste superior ao obtido pela log normal para as três cidades estudadas (Bela Vista do Paraíso, Ibipora e Marilândia do Sul).

**Palavras-chave:** *Gama, Índice Padronizado de Precipitação, Log normal.*

---

<sup>1</sup>DECEB – UFSJ / CSL : Universidade Federal de São João del Rei – *Campus Sete Lagoas*, MG. Email: moreira.raissa@hotmail.com

<sup>2</sup> DECEB – UFSJ / CSL : Universidade Federal de São João del Rei – *Campus Sete Lagoas*, MG. Email:anapaula@ufsj.edu.br

<sup>3</sup> DEPEN – IFBA: Instituto Federal da Bahia, Ba. Email: jailsondearaujo@yahoo.com.br

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

## **Estudo da relação entre os tipos de deficiência e variáveis socioeconômicas da região do Triângulo Mineiro e Alto Paranaíba**

**Bruna Queiroz de Melo Prado<sup>1 4</sup>**

**Nádia Giaretta Biase<sup>2 4</sup>**

**Tiago Bortoletto Lino dos Santos<sup>3 4</sup>**

Ao longo da história sempre existiram pessoas com alguma limitação física, sensorial ou intelectual, cuja trajetória foi marcada pela ignorância, indiferença, preconceito e lutas em busca da inclusão social. No Brasil, o censo de 2000 registrou 24,5 milhões de pessoas com algum tipo de deficiência. A avaliação dos deficientes de uma região permite conhecer melhor as condições e realidades nas quais essas pessoas estão inseridas. São poucos os estudos voltados às pessoas com deficiência, especificamente em relação à qualidade de vida dessas pessoas nas regiões do Triângulo Mineiro e Alto Paranaíba de Minas Gerais. De acordo com os resultados, por meio da análise descritiva dos dados obtidos a partir do censo demográfico de 2010 (IBGE), tem-se que o tipo de deficiência mais frequente entre os habitantes residentes da região do Triângulo Mineiro e Alto Paranaíba é a deficiência visual. Ambas as regiões não apresentaram diferença significativa entre si em relação ao tipo de deficiência e as variáveis socioeconômicas domicílio, sexo e renda, porém através da aplicação do teste Qui-quadrado verificou-se a associação entre os tipos de deficiência e as variáveis socioeconômicas.

**Palavras-chave:** Minas Gerais; análise descritiva; Qui-quadrado.

---

<sup>1</sup> Faculdade de Matemática da Universidade Federal de Uberlândia, Uberlândia-MG. Email: bruna.est.ufu@hotmail.com

<sup>2</sup> Faculdade de Matemática da Universidade Federal de Uberlândia, Uberlândia-MG. Email: nadiabiase@gmail.com

<sup>3</sup> Faculdade de Matemática da Universidade Federal de Uberlândia, Uberlândia-MG. Email: tiago.bortoletto@hotmail.com

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

# Aplicação dos modelos de Holt-Winters aos dados de produção de leite do estado do Paraná (PR)

Mayra Marques Bandeira<sup>1</sup>

Eduardo Campana Barbosa<sup>2</sup>

Rômulo César Manuli<sup>3</sup>

Filipe Ribeiro Formiga Teixeira<sup>4</sup>

Matheus De Paula Ferreira<sup>5</sup>

Explica-se o aumento da produção de leite no Brasil principalmente pela reestruturação do setor lácteo a partir de 1990, além da recente busca por mercados externos. Estes acontecimentos priorizaram o planejamento do seu processo produtivo, enfatizando o quão importante é o estudo das tendências deste mercado. Neste trabalho analisou-se à série histórica de produção de litros de leite do estado do Paraná (PR), composta por observações mensais de 10 anos (jan./2003 a dez./2012), no intuito de fazer previsão para os 9 primeiros meses de 2013. Aplicou-se os modelos de Holt-Winters aditivo e multiplicativo, implementados no pacote *forecast* do *software* R. A estimação dos parâmetros (constantes de suavização) visou a minimização da soma dos quadrados dos erros. Para selecionar o melhor modelo, utilizou-se o indicador de erro MAPE (Erro Percentual Médio Absoluto). A análise residual consistiu em aplicar os testes Box-Pierce, Shapiro-Wilk e correlogramas dos resíduos quadráticos, no intuito de verificar, respectivamente, a independência, normalidade e homogeneidade de variância dos erros. O modelo aditivo de Holt-Winters foi escolhido por apresentar menor MAPE (3,84%) em relação ao multiplicativo (5,72%). Além disso, os resíduos deste modelo atenderam as pressuposições citadas, com propriedades de ruído branco, o que corrobora a qualidade do ajuste. As estimativas dos parâmetros deste modelo foram:  $\alpha = 0,7282$  (nível),  $\beta = 0,0097$  (tendência) e  $\gamma = 0,6490$  (sazonalidade).

**Palavras-chave:** *séries temporais, previsão, ruído branco.*

---

<sup>1</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [mayra.mbandeira@gmail.com](mailto:mayra.mbandeira@gmail.com)

<sup>2</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [duducampana@hotmail.com](mailto:duducampana@hotmail.com)

<sup>3</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [romulomanuli@ig.com.br](mailto:romulomanuli@ig.com.br)

<sup>4</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [filipeformiga1@live.com](mailto:filipeformiga1@live.com)

<sup>5</sup> DET-UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [matheusferreiraufv@gmail.com](mailto:matheusferreiraufv@gmail.com)

# Modelagem de Arrecadação de Prêmios de Seguro de Automóveis do Rio de Janeiro

Cíntia Yuki Teodoro Ogawa<sup>1 7 8</sup>; Rafael Agostinho Ferreira<sup>2 7 8</sup>; Nayane Humberto Ferreira<sup>3 7 8</sup>; Paulo Henrique Medeiros Bragança<sup>3 7 8</sup>; Leandro Ferreira<sup>5 7 8</sup>; Marçal Serafim Cândido<sup>6 7 8</sup>.

No mercado securitário automobilístico, se torna cada vez mais desafiador estimar os riscos para os produtos atuariais ofertados, uma vez que as variáveis que influenciam no seu cálculo estão cada vez mais complexas. No objetivo de auxiliar o cálculo dos riscos atuariais, busca-se neste trabalho, aplicar a metodologia de Modelos Lineares Generalizados (MLG), a fim de criar um modelo capaz de apresentar uma previsão satisfatória para os prêmios de seguro de automóveis no estado do Rio de Janeiro. Para tal, utilizou-se como variáveis que explicam o valor do prêmio o índice de sinistralidade, o número total de licenciamentos de veículos no estado do Rio de Janeiro, o índice de preços IPCA, a taxa de juros SELIC e a taxa de recuperabilidade de veículos (razão entre a quantidade de veículos recuperados e quantidade de veículos roubados ou furtados), e o rendimento médio real do trabalhador.

**Palavras-chave:** Modelos lineares generalizados, Seguro, Automóveis.

---

<sup>1</sup>Email: [cinyuki@gmail.com](mailto:cinyuki@gmail.com)

<sup>2</sup>Email: [rafaelferreiratc@hotmail.com](mailto:rafaelferreiratc@hotmail.com)

<sup>3</sup>Email: [nayanemikaity@hotmail.com](mailto:nayanemikaity@hotmail.com)

<sup>3</sup>Email: [braganca\\_paulo@live.com](mailto:braganca_paulo@live.com)

<sup>5</sup>Email: [leandro.ferreira@unifal-mg.edu.br](mailto:leandro.ferreira@unifal-mg.edu.br)

<sup>6</sup>Email: [marcal.serafim@unifal-mg.edu.br](mailto:marcal.serafim@unifal-mg.edu.br)

<sup>7</sup>UNIFAL-MG: Universidade Federal de Alfenas - Minas Gerais.

<sup>8</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

# DEPENDÊNCIA ESPACIAL DO IDEB-2011 PARA O 5º ANO DO ENSINO FUNDAMENTAL E DO IDH-2010 PARA A MESORREGIÃO DO TRIÂNGULO MINEIRO E ALTO PARANAÍBA.

LUIZ CARLOS COSTA JÚNIOR<sup>1,4</sup>

VIVIAN RIBEIRO BARRETO<sup>3,4</sup>

EDNALDO CARVALHO GUIMARÃES<sup>2,4</sup>

O objetivo foi estudar os índices do IDEB 2011 para o 5º ano e IDH 2010 para a região do Triângulo Mineiro e Alto Paranaíba, que totalizam 66 cidades. Foram analisados os dados das avaliações obtidos pelos *sites* do Ministério da Educação e do Programa das Nações Unidas para o Desenvolvimento. A partir dos dados obtidos pelos *sites*, foi feita a geo-referênciação das cidades por meio de latitude e longitude. Para a análise espacial foi utilizada a função semivariância, em seguida foram ajustados os modelos teóricos (semivariâncias experimentais), e definidos os parâmetros dos semivariogramas (Patamar, Alcance, Efeito Pepita) e para a criação dos mapas foi utilizada a interpolação por meio da krigagem ordinária. Em toda região estudada, os valores do IDEB foram abaixo de 6, nota ideal, porém a região do Alto Paranaíba obteve uma melhor nota que o Triângulo Mineiro. Conclui-se pela a análise, que o Triângulo Mineiro possui um índice menor que o Alto Paranaíba nos temas analisados. Porém ambas as regiões devem intensificar mais os trabalhos na educação, para que em futuras avaliações, o índice ideal seja igualado ou superado.

**Palavras chave:** IDEB, IDH, geoestatística.

- 1- Acadêmico do curso de Bacharelado em Estatística FAMAT /UFU – Bolsista PIBIC/CNPQ (agosto/13 a julho/14) - Av. João Naves de Ávila, 2160, Bairro Santa Mônica, Uberlândia – MG, CEP: 38400-900 – [luizcarloscostajr11@gmail.com](mailto:luizcarloscostajr11@gmail.com).
- 2- Prof. Orientador – FAMAT/UFU [ecg@ufu.br](mailto:ecg@ufu.br)
- 3- Acadêmico do curso de bacharelado em Estatística FAMAT / UFU – Av. João Naves de Ávila, 2160, Bairro: Santa Mônica, Uberlândia – MG, CEP: 38400-900 – [vivian\\_r\\_barreto@hotmail.com](mailto:vivian_r_barreto@hotmail.com).
- 4- Agradecimentos a FAPEMIG pelo apoio financeiro e bolsa iniciação científica do aluno

# Visão política dos universitários

Aline Soier Sanches<sup>1 4</sup>

Diogo Maia de Figueiredo<sup>2 4</sup>

Raquel Yuri da Silveira Aoki<sup>3 4</sup>

Motivados pela repercussão das manifestações ocorridas em junho e julho de 2013, surgiu o interesse em pesquisar a visão política dos universitários; uma vez que estes representariam maior porcentagem dos participantes das manifestações, e pesquisar sua visão sobre esses acontecimentos. Foi notável o caráter difuso das manifestações e o objetivo foi traçar o perfil político, e os conhecimentos gerais dos universitários sobre o conteúdo desses movimentos. A pesquisa foi realizada na UFMG e na PUC (unidade Coração Eucarístico). Primeiro o estudante abordado respondia as questões de conhecimento específico e em seguida a sua visão sobre as manifestações conforme havia sido a sua participação. Foi aplicada uma amostra piloto igual a 100 questionários proporcional ao número de alunos matriculados nas universidades em questão, segundo informações obtidas no ano de 2010. Através do Modelo Rasch foi obtido a variabilidade do conhecimento político dos universitários entrevistados e através deste foi calculado o tamanho da amostra usando amostragem estratificada com alocação ótima. Foi observado como resultados que não há diferença entre os níveis de conhecimento político dos estudantes da UFMG e da PUC- Coração Eucarístico.

**Palavras-chave:** *Política, Estatística, Amostragem, Manifestação, Universitários.*

---

<sup>1</sup> Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [alinesoier93@gmail.com](mailto:alinesoier93@gmail.com)

<sup>2</sup> Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [diogo.mfigueiredo63@gmail.com](mailto:diogo.mfigueiredo63@gmail.com)

<sup>3</sup> Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [raquel.aoki@gmail.com](mailto:raquel.aoki@gmail.com)

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

# **Análise Fatorial de doenças relacionadas ao saneamento inadequado no Brasil entre 1993 a 2010**

**Bruna Queiroz de Melo Prado<sup>1 4</sup>**

**Priscila Neves Faria<sup>2 4</sup>**

**Edilamar Queiroz de Melo<sup>3 4</sup>**

Estudos vêm sendo desenvolvidos a partir do início da década de 80, buscando formular mais rigorosamente os mecanismos responsáveis pelo comprometimento das condições de saúde da população, na ausência de condições adequadas de saneamento. Neste sentido, o presente trabalho teve como objetivo utilizar a Análise Fatorial no estudo das incidências de Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI). As análises foram feitas por meio do número de internações hospitalares anuais no Brasil, entre o período de 1993 a 2010. Através da Análise Fatorial dois fatores foram extraídos na análise, que em conjunto foram capazes de explicar 81% da variância total dos dados além de destacar, dentre as variáveis estudadas, as que mais influenciaram nas questões voltadas ao saneamento básico. Além disso, o fator 1 foi representado por doenças relacionadas a condições sanitárias e o fator 2 por doenças transmitidas por insetos vetores.

**Palavras-chave:** condições sanitárias, insetos vetores, incidência.

---

<sup>1</sup> Faculdade de Matemática da Universidade Federal de Uberlândia, Uberlândia-MG. Email: bruna.est.ufu@hotmail.com

<sup>2</sup> Faculdade de Matemática da Universidade Federal de Uberlândia, Uberlândia-MG. Email: priscila@famat.ufu.br

<sup>3</sup> Instituto de Geografia da Universidade Federal de Uberlândia, Uberlândia-MG. Email: edilamar.queiroz@bol.com.br

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

# Análise de Dados Amostrais Complexos em Inquéritos Populacionais de Saúde

Dário Alves da Silva Costa<sup>1 6</sup>

Priscila Brandão dos Reis<sup>2 6</sup>

Amanda Cristina de Souza Andrade<sup>3 6</sup>

Amélia Augusta de Lima Friche<sup>4 6</sup>

Waleska Teixeira Caiaffa<sup>5 6</sup>

Inquéritos são importantes para avaliar o estado da saúde da população e demandam métodos de amostragens complexos. Este estudo objetiva avaliar o efeito do desenho em amostragem complexa nas estimativas dos parâmetros de interesse e seus níveis de precisão. Foram utilizados os dados do inquérito domiciliar de residentes em dois distritos sanitários de Belo Horizonte, realizado pelo Observatório de Saúde Urbana de Belo Horizonte (2008-2009). A amostra, obtida a partir de um delineamento probabilístico estratificado por conglomerados em três estágios: (a) setor censitário, (b) domicílio, (c) indivíduo, considerou como fator de estratificação um indicador complexo de vulnerabilidade composto por saneamento, habitação, educação, renda, social e saúde. O impacto da incorporação do efeito do plano amostral (EPA) foi avaliado pela medida DEFF (design-effect). As maiores variações encontradas foram na variável escolaridade: 0,84 (analfabetos) a 5,00 (indivíduos com 12 anos ou mais de estudo). Houve redução na prevalência da Hipertensão (30,83% para 23,36%) e Diabetes (8,04% para 6,08%), que apresentaram DEFF de 1,54 e 1,17, respectivamente. A associação entre sexo e hipertensão, ignorando o plano amostral, apresentou um  $p=0,004$  e considerando o desenho foi de  $p=0,112$ . Análises que não incorporam características do desenho amostral podem produzir incorreções nas estimativas médias e erros padrões, comprometendo os resultados, os testes de hipóteses e as conclusões da pesquisa.

**Palavras-chave:** *Amostragem, Desenho Complexo, Inquérito de Saúde.*

---

<sup>1</sup> Mestrando Saúde Pública/Faculdade de Medicina – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte - MG. Email: darioalves\_sc@yahoo.com.br

<sup>2</sup> Mestranda Saúde da Criança e Adolescente/Faculdade de Medicina – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte - MG. Email: pribreis@gmail.com

<sup>3</sup> Doutoranda Saúde Pública/Faculdade de Medicina – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte - MG. Email: amandasouza\_est@yahoo.com.br

<sup>4</sup> Professora adjunta Fonoaudiologia/Faculdade de Medicina – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte - MG. Email: gutafriche@gmail.com

<sup>5</sup> Professora Titular Medicina/Faculdade de Medicina – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte - MG. Email: wcaiaffa@medicina.ufmg.br

<sup>6</sup> Agradecimento à FAPEMIG pelo apoio financeiro.

## Tratamento multivariado de dados da População Economicamente Ativa no Brasil via Análise de Agrupamentos

Laíla Luana Campos<sup>1</sup>  
Priscila Neves Faria<sup>2</sup>

A População Economicamente Ativa (PEA) é um conceito elaborado para designar a população que está inserida no mercado de trabalho ou que, de certa forma, está procurando se inserir nele para exercer algum tipo de atividade remunerada. Assim, a parte da população que está desempregada e que não busca empregos, como crianças menores que 10 anos, estudantes que não trabalham, donas de casa que exercem apenas funções domésticas não remuneradas, entre outros, é incluída naquilo que se denomina por População Economicamente Inativa. A denominação “Análise Multivariada” corresponde a um conjunto de métodos e técnicas que analisam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados. Dentro desta análise, tem-se a técnica de Análise de Agrupamentos, cujo objetivo principal é situar as observações homogêneas em grupos, a fim de definir uma estrutura para os dados, agrupando os elementos da amostra ou população em grupos de forma que elementos de um mesmo grupo são homogêneos entre si, no que se refere às variáveis (características) que neles foram medidas. Por outro lado estes grupos já formados são heterogêneos entre eles em relação a estas mesmas características. Como resultado, foi obtido que, em ambos os sexos, os gráficos dendrogramas referente à PEA nas regiões metropolitanas brasileiras pesquisadas foram similares, uma vez que em ambos foram formados 5 grupos, e nestes, as mesmas regiões os compunham em cada gráfico.

**Palavras-chave:** similaridade, agrupamentos, mercado de trabalho.

---

<sup>1</sup> Universidade Federal de Uberlândia – Faculdade de Matemática, Uberlândia, MG. E-mail: lailaluanacampos@gmail.com

<sup>2</sup> Universidade Federal de Uberlândia – Faculdade de Matemática, Uberlândia, MG. E-mail: priscila@famat.ufu.br

## **Análise estatística de incidentes de segurança em servidores de e-mail no Brasil : Ocorrência de Spams no ano de 2013.**

Fernanda Ires Silva<sup>1</sup>

José Francisco Clementino de Jesus<sup>2</sup>

Bárbara de Oliveira Aguiar<sup>3</sup>

Maria de Fátima Ferreira Almeida<sup>4</sup>

Enviar uma mensagem, geralmente de fins comerciais, para vários endereços eletrônicos ao mesmo tempo, denomina-se Spam (*Sending and Posting Advertisement in Mass*). Disseminar panfletos comerciais através dos servidores de e-mails, tornou-se financeiramente mais viável, do que investir em propagandas em canais de Televisão, ligações telefônicas entre outros. Porém, esta ação passou a ser usada para fins negativos, tornando, desta forma, um grande problema para usuários de e-mails. Através de mensagens chamativas o usuário é induzido a clicar em um determinado anúncio que pode ser simplesmente um *link* para o incômodo. Diante deste cenário, o estudo propõe um análise estatística a respeito das incidências de *Spams* reportados ao Cert (Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil) no espaço amostral brasileiro no período de janeiro a dezembro de 2013. O Teorema de Bayes, uma das relações mais importantes sobre probabilidade (MORETTIN, 2010), possibilita uma análise da probabilidade de ocorrência de *Spam* dentre os incidentes de segurança reportados no Brasil. De posse dos dados fornecidos pelo Cert e o tratamento probabilístico, é possível concluir o quanto predomina a disseminação de *Spams* sobre os demais incidentes de segurança em servidores de e-mails brasileiros e o quanto se tem que avançar na segurança da internet.

**Palavras-chave:** *Probabilidade, Spam, e-mail.*

---

1 IFNMG- *campus* Montes Claros: Instituto Federal de Educação Ciência e Tecnologia do Norte de Minas Gerais. Email: iresnanda@gmail.com

2 IFNMG- *campus* Montes Claros: Instituto Federal de Educação Ciência e Tecnologia do Norte de Minas Gerais. Email: rj45francis@gmail.com

3 IFNMG- *campus* Montes Claros: Instituto Federal de Educação Ciência e Tecnologia do Norte de Minas Gerais. Email: barbara.oliveira.aguiar@gmail.com

4 IFNMG- *campus* Montes Claros: Instituto Federal de Educação Ciência e Tecnologia do Norte de Minas Gerais

# Estudo do perfil dos candidatos do ENEM em 2011 por meio de inferência não-paramétrica no município de Diamantina-MG

Thacyo Euqueres De Villa<sup>1 5</sup>

João Cláudio Quirino Cardoso<sup>2 5</sup>

Douglas Silva Oliveira<sup>3 5</sup>

Janser Moura Pereira<sup>4 5</sup>

O presente trabalho tem como objetivo verificar se há diferença significativa no rendimento escolar dos candidatos por meio das notas do Exame Nacional do Ensino Médio (ENEM), em relação à dependência administrativa das escolas e em função do gênero, no município de Diamantina-MG. Para a pesquisa foram utilizados dados do ENEM, obtidos a partir do portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), referentes ao ano de 2011. Uma vez que a pressuposição de normalidade não foi atendida, aplicou-se o teste não-paramétrico de Kruskal-Wallis, ao nível de significância de 5%, conclui-se que há diferença significativa entre as medianas das notas em relação ao sexo do candidato e a dependência administrativa da escola em que estudam.

**Palavras-chave:** *Rendimento escolar, Teste não-paramétrico, Kruskal-Wallis.*

---

<sup>1</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: [thacyo@est.ufu.br](mailto:thacyo@est.ufu.br)

<sup>2</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: [joaoclaudioq@gmail.com](mailto:joaoclaudioq@gmail.com)

<sup>3</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: [douglasso1988@gmail.com](mailto:douglasso1988@gmail.com)

<sup>4</sup>FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia-MG.  
Email: [janser@famat.ufu.br](mailto:janser@famat.ufu.br)

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

# **Ajuste de distribuições de probabilidade para a depreciação das instalações na produção de suínos para o abate, Rio Grande do Sul, durante os anos de 2006 a 2013**

**Laila Luana Campos<sup>1</sup>**  
**Tatiane Bento da Costa<sup>2</sup>**  
**Marcelo Tavares<sup>3</sup>**

Os Bens Patrimoniais, devido ao uso e pelo desgaste natural ao longo do tempo, sofrem desvalorização de seu valor original, a qual se denomina “depreciação”. Na análise da composição total de custos de produção de suínos para o abate, esta é uma variável extremamente importante. Neste trabalho objetivou-se buscar um bom ajustamento das distribuições de probabilidade na variável depreciação das instalações, a qual é uma das variáveis de custo fixo na produção de suínos para o abate, nos meses de janeiro a junho (meses em que os dados são fornecidos), durante os anos de 2006 a 2013, no estado do Rio Grande do Sul. Os dados foram retirados do site da CONAB (Companhia Nacional de Abastecimento) e ajustados às funções de distribuições Weibull, Rayleigh, Lognormal e Normal, mensalmente. Considerando um nível de significância de 5%, verifica-se que todas as distribuições de probabilidade analisadas ajustam-se à depreciação das instalações.

**Palavras chaves:** custo, suínos, distribuição de probabilidades.

---

<sup>1</sup> Universidade Federal de Uberlândia – Faculdade de Matemática, Uberlândia, MG. E-mail: lailuanacampos@gmail.com

<sup>2</sup> Universidade Federal de Uberlândia – Faculdade de Administração, Uberlândia, MG. E-mail: tatianecosta@adm.ufu.br

<sup>3</sup> Universidade Federal de Uberlândia – Faculdade de Matemática, Uberlândia, MG. E-mail: mtavares@famat.ufu.br

# Análise de Sobrevivência: Um comparativo entre os *softwares* SAS e R

Brandon Ferreira de Mendonça <sup>1 2</sup>

A análise de dados do tipo tempo até a ocorrência de um evento de interesse, geralmente denominada por análise de sobrevivência, surge em diversas áreas do conhecimento, tais como medicina, biologia, engenharias, saúde pública, epidemiologia, ciências sociais e economia. Dados de sobrevivência são frequentemente caracterizados pela presença de informações incompletas, conhecidas como censuras, e por este motivo, requerem técnicas de análise especiais para acomodar tais informações incompletas. Existem atualmente diversos *softwares* estatísticos disponíveis para a análise de dados de sobrevivência, dentre os quais destacam-se os *softwares* SAS e R. O *software* SAS é, muito provavelmente, o *software* comercial mais utilizado pela comunidade estatística em geral, enquanto o *software* R, que é gratuito, tem sido amplamente utilizado pela comunidade acadêmica, sendo, portanto, de grande importância para o analista saber como realizar análises utilizando tais *softwares*. Neste contexto, o principal objetivo desse trabalho é a realização de um estudo comparativo envolvendo os *softwares* SAS e R, para execução das técnicas mais usuais de análise de dados de sobrevivência.

**Palavras-chave:** *Métodos não-paramétricos, modelos paramétricos, modelos de riscos proporcionais.*

---

<sup>1</sup>DEST-UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Email: [brandon@ufmg.br](mailto:brandon@ufmg.br)

<sup>2</sup>Agradecimento ao prof. Fábio Nogueira Demarqui (DEST-UFMG) pela orientação e à FAPEMIG pelo apoio financeiro.

## **Relações da Empregabilidade por Setor de Atividade Econômica no Estado de Minas Gerais no Ano de 2013**

**Paulo César Ossani**<sup>1</sup>

**Marcelo Ângelo Cirillo**<sup>2</sup>

Os Setores da Atividade Econômica empregam milhões de pessoas, a fim de investigar as relações de empregabilidade no estado de Minas Gerais em 2013, dados com o saldo de Admissões-Demissões foram extraídos do site do Ministério do Trabalho. Usando a Análise de Componentes Principais para observar as variáveis correlacionadas, verificou-se que o setor *Comércio* correlaciona-se forte com o setor *Serviços*, e o setor *Agropecuário* possui uma correlação inversa com os outros setores, confrontando esses resultados com as classificações da Análise de Clusters, utilizando a métrica Euclidiana, concluímos que os setores *Comércio* e *Serviços* possuem alta correlação e alta similaridade na empregabilidade, e os outros setores possuem baixa correlação e baixa similaridade. Como o setor *Agropecuário* possui correlação inversa com os outros setores, sugere que no ano de 2013 o saldo de empregos nesse setor foi refletido diretamente pelos outros setores.

**Palavras-chave:** *Empregos, Minas Gerais, Correlações, PCA, Análise de Clusters.*

---

<sup>1</sup> Mestrando em Est. Exp. Agrop., Depto Ciências Exatas, UFLA, Email: [ossanipc@hotmail.com](mailto:ossanipc@hotmail.com)

<sup>2</sup> Prof. Adjunto IV, Depto Ciências Exatas, UFLA, Email: [macufla@dex.ufla.br](mailto:macufla@dex.ufla.br)

# Análise da evolução temporal do preço médio da carne em Belo Horizonte, MG.

Tamara Aparecida Nogueira dos Anjos<sup>1 5</sup>

Érica Fernanda da Cruz<sup>2 5</sup>

Thelma Safadi<sup>3 5</sup>

Adriano Carvalho Costa<sup>4 5</sup>

A pesquisa da Cesta Básica Nacional, também denominada Ração Essencial Mínima, realizada pelo Departamento Intersindical de Estatísticas e Estudos Socioeconômicos (DIEESE), em dezesseis capitais do Brasil, acompanha mensalmente a evolução de preços de treze produtos de alimentação. Dentre os produtos que compõem a Cesta Básica, destaca-se a carne, que é uma importante fonte de proteína e deve fazer parte de uma dieta balanceada. Neste trabalho buscou-se analisar a evolução temporal do preço médio da carne em Belo Horizonte, MG, no período de agosto de 1994 a junho de 2014, utilizando técnicas de Séries Temporais. Observou-se uma tendência crescente na série descrita por  $\hat{T}_t = 1,6344 + 0,0583t$ , obtida pelo método dos mínimos quadrados e uma sazonalidade de 12 meses. Para análise dos preços utilizou-se o modelo SARIMA, com o intuito de descrever o comportamento da série e encontrar um modelo que forneça melhores previsões. A adequação do modelo foi verificada através do teste Box-Pierce e escolheu-se aquele que possui melhor ajuste pelo critério de Informação Akaike (AIC). O modelo escolhido foi o SARIMA(2, 1, 2)X(0, 1, 1)<sub>12</sub>, o qual apresentou projeções dos preços satisfatórias.

**Palavras-chave:** *Carne, Série Temporal, Tendência, Sazonalidade.*

---

<sup>1</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [tamarananjos@gmail.com](mailto:tamarananjos@gmail.com)

<sup>2</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [erica.89@bol.com.br](mailto:erica.89@bol.com.br)

<sup>3</sup>DEX-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [safadi@dex.ufla.br](mailto:safadi@dex.ufla.br)

<sup>4</sup>DZO-UFLA: Universidade Federal de Lavras, Lavras-MG. Email: [acarvalhocosta@gmail.com](mailto:acarvalhocosta@gmail.com)

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

# Modelagem Inteligente para Análise de Séries de Vazões e Afluentes

Bethânia Oliveira de Brito <sup>1 4</sup>

Luiz Alberto Beijo <sup>2 4</sup>

Ricardo Menezes Salgado <sup>3 4</sup>

As usinas hidrelétricas são responsáveis pela maior parte do abastecimento da demanda de energia elétrica no Brasil. Portanto, é de extrema importância conhecer bem o comportamento dos fenômenos hídricos que são bastante complexos. Surge então, a necessidade de estabelecer modelos que sejam capazes de representar de forma eficiente o comportamento da água em determinada usina. De forma geral, estudos desse tipo são úteis no planejamento da operação de uma usina hidrelétrica. Existem modelos que são propostos por diversos autores para modelar as séries de vazões tais como, os modelos inteligentes como o RNAs, modelos de redes neurais nebulosas e os modelos tradicionais de séries temporais como os modelos bayesianos. Porém, a conclusão sobre qual o melhor modelo ainda é bastante contraditória. Propôs-se nesse trabalho a utilização de um combinador, através do qual foi possível unir os benefícios do modelo inteligente e do modelo bayesiano, obtendo assim estimativas mais precisas. O combinador utilizado neste trabalho foi composto utilizando dois tipos de modelos: RNAs e Modelos Bayesianos. A estratégia de combinação mostrou-se adequada ao problema provendo resultados competitivos em relação aos modelos individuais. Nesta óptica, pode-se concluir que a técnica proposta foi adequada ao problema proposto, podendo ser utilizada no contexto de planejamento do setor elétrico.

**Palavras-chave:** *Séries Temporais, Modelos Inteligentes, Séries de Vazões.*

## Referências

- [1] BALLINI, R.. *Análise e previsão de séries de vazões utilizando modelos de séries temporais, redes neurais e redes neurais nebulosas*. 169f. Tese (Doutorado em engenharia elétrica)- Faculdade de engenharia elétrica e computação, Universidade Federal de Campinas, Campinas, 2000.
- [2] SALGADO, R. M.; PEREIRA, J. J. F. et al . *A Hybrid Ensemble Model Applied to the Short-Term Load Forecasting Problem*. In: International Joint Conference on Neural Networks, 2006 Vancouver, BC, Canada, 2006.

---

<sup>1</sup>UNIFAL - Universidade Federal de Alfenas, Alfenas-MG. Email: *brito.bethania@gmail.com*

<sup>2</sup>UNIFAL - Universidade Federal de Alfenas, Alfenas-MG. Email: *luiz.beijo@unifal-mg.edu.br*

<sup>3</sup>UNIFAL - Universidade Federal de Alfenas, Alfenas-MG. Email: *ricardo@bcc.unifal-mg.edu.br*

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

# Simulação em R para Análise de Desempenho de Gráficos de Controle de $\bar{X}$

Mário Halfeld Clark Campos<sup>1</sup>

Lupércio França Bessegato<sup>2</sup>

Os gráficos de controle de Shewart são uma das principais ferramentas do controle estatístico de qualidade. Sendo a característica expressa numericamente em escala contínua de medidas, o usual é monitorar o processo por um par de gráficos de controle para variáveis: um para monitorar a centralidade e outro, a dispersão. Os mais usuais são: o gráfico de controle para média amostral ( $\bar{X}$ ) e para amplitude amostral (R).

Nesse trabalho, apresenta-se uma rotina em R em que são feitas simulações de Monte Carlo em situações de o processo estar sob controle e fora de controle. A rotina desenvolvida permite verificar o desempenho de carta de controle ( $\bar{X}$ ) para qualquer distribuição subjacente de característica de qualidade contínua.

Desenvolvemos script para construção de gráficos das simulações, na linguagem do programa estatístico R (R Core Team, 2013) utilizando os pacotes *tcltk*; e *TeachingDemos*. Assim, pôde-se implementar janela interativa, amigável com o usuário, que facilita a manipulação dos valores de entrada da simulação.

Os resultados deste estudo por simulação indicam a utilidade desta rotina na avaliação das taxas de alarmes falsos e poder de gráficos de  $\bar{X}$  em processos sob controle e fora de controle.

**Palavras-chave:** Gráfico de  $\bar{X}$ , simulação de Monte Carlo, taxa de alarme falso, poder, comprimento médio de sequência e janela interativa amigável com o usuário.

---

<sup>1</sup> DE/ICE – UFJF. e-mail: marioclark07@yahoo.com.br.

<sup>2</sup> DE/ICE – UFJF. e-mail: lupercio.bessegato@ufjf.edu.br.

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

# Efeito da Transformação de Dados na Análise de Componentes Principais Aplicada ao Estudo da Requeima da Batata

Matheus de Paula Ferreira<sup>1 5</sup>

Rafael Botelho Barbosa<sup>2 5</sup>

Ana Carolina Campana Nascimento<sup>3 5</sup>

Moisés Nascimento<sup>4 5</sup>

A batata é um dos alimentos mais consumidos no mundo, porém sua produção tem sido limitada pela infecção de doenças como a requeima, que é uma doença causada pelo fungo *Phytophthora infestans*, resultando em necrose nas folhas, ramos e tubérculos, deixando a planta infectada com aspecto queimado. Dito isto, tem-se que o conhecimento de cultivares menos susceptíveis à esta doença torna-se importante. Uma forma de realizar esta identificação se dá a partir da aplicação da análise de componentes principais (CP). A análise de CP é uma técnica de estatística multivariada para modelar a estrutura da matriz de covariâncias das variáveis, para tanto, geralmente usa-se a matriz de covariância (S), correlação (R) ou outras abordagens, como a proposta por CAMPANA et al. (2010). Deste modo, este trabalho teve por objetivo identificar os cultivares mais resistentes a partir dos resultados obtidos pela análise de CP e ainda, comparar os *rankings* de cultivares a partir das três abordagens supracitadas. Os dados são provenientes de um experimento em DBC no município de Maria da Fé, Minas Gerais, de 22 de março a 8 de julho de 2008. No total foram avaliados 34 cultivares e 6 variáveis. A análise de CP permitiu reduzir a dimensionalidade dos dados, sendo necessário apenas um componente principal. Ao compararmos os resultados obtidos a partir das três abordagens observamos que os *rankings* não apresentaram diferença, mesmo quando apresentaram valores distintos para os escores na técnica de CP.

**Palavras-chave:** *Análise Multivariada, Componentes Principais, Requeima.*

---

<sup>1</sup> UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [matheusferreiraufv@gmail.com](mailto:matheusferreiraufv@gmail.com)

<sup>2</sup> UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [rafael.botelho@ufv.br](mailto:rafael.botelho@ufv.br)

<sup>3</sup> UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [campanaac@gmail.com](mailto:campanaac@gmail.com)

<sup>4</sup> UFV: Universidade Federal de Viçosa, Viçosa-MG. Email: [moysesnascim@ufv.br](mailto:moysesnascim@ufv.br)

<sup>5</sup> Agradecimento à CAPES, CNPQ e FAPEMIG pelo apoio financeiro.

# Normalidade em modelos mistos usando um método de transformação nos dados

Natália da Silva Martins<sup>1 2</sup>

Alessandra dos Santos<sup>3 4</sup>

A inferência estatística, frequentemente assume normalidade dos resíduos do modelo em estudo. A pressuposição de normalidade pode ser verificada por métodos gráficos ou testes estatísticos, aos quais são aplicáveis quando há independência residual. Em modelos lineares mistos existem dois tipos de resíduos, marginal e condicional, e estes não são independentes. Com isso, foram propostos os métodos gráficos de probabilidade normal para uma modificação do preditor EBLUP, probabilidade normal para os resíduos condicionais padronizados e de probabilidade normal para os resíduos de confundimento mínimo. Para realização de testes, por haver a correlação, os métodos multivariados são indicados para avaliar a normalidade das observações. No entanto, estes nem sempre detectam corretamente a estrutura dos dados. Pensando nestas situações, Xu, Li e Song (2013), apresentaram um estudo de dados simulados, admitindo um modelo aleatório de dois fatores, com e sem interação. Estes autores propõem uma transformação nos dados, de tal forma que este pondere a estrutura de correlação existente e na realização dos testes multivariados de Royston, assimetria (raiz b1) e MQ, sobre suposições de diferentes distribuições para o erro, com tamanhos amostrais e parâmetros de dispersão, estes melhoram o poder de rejeição dos testes de normalidade. Deste modo, este trabalho analisou a ideia dos autores e repetiu o estudo de simulação, verificando que a proposta é viável e interessante, pois aumenta o poder de rejeição conforme aumenta o tamanho amostral, sendo que os dados ao receberem a transformação indicam uma maior precisão.

**Palavras-chave:** *dados correlacionados, modelos mistos, normalidade residual, poder do teste.*

<sup>1</sup>ICEEx-Unifal: Universidade Federal de Alfenas, Alfenas-MG. Email: [natalia.martins@unifal-mg.edu.br](mailto:natalia.martins@unifal-mg.edu.br)

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

<sup>3</sup>LCE-ESALQ/USP: Universidade de São Paulo, Piracicaba-SP. Email: [alessandra.s@usp.br](mailto:alessandra.s@usp.br)

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

# Predição do lucro de empresas brasileiras de capitalização utilizando redes neurais artificiais: um estudo introdutório

Richard Hernani Pereira<sup>1 4</sup>

Leandro Ferreira<sup>2 4</sup>

Marçal Serafim Candido<sup>3 4</sup>

Apresenta-se um estudo introdutório com objetivo de prever o lucro das dez maiores empresas brasileiras de capitalização, utilizando modelos baseados em redes neurais artificiais através de indicadores econômicos, tais como: Produto Interno Bruto, Índice Geral de Preços do Mercado, Taxa Referencial de Juros e o Índice Bovespa. As RNA's são modelos computacionais estatísticos, considerados sistemas inteligentes e inspiram-se nos mecanismos de aprendizagem do cérebro humano. Foram coletados dados mensais de 11/2003 a 11/2013, considerando dois conjuntos de dados, sendo o conjunto A para treinamento (60%) e teste (40%) e o conjunto B para predição do lucro após o treinamento das redes. Executou-se três modelos de arquiteturas diferentes; arquiteturas: (1) 4 neurônios de entrada, 5 na camada escondida e 1 neurônio na saída; (2) 4 neurônios de entrada; 10 na camada escondida e 1 na saída; (3) 4 neurônios na entrada, 15 na camada escondida e 1 na saída, através do software *FannTool* 1.0. As estimativas obtidas são próximas para as diferentes arquiteturas. A arquitetura 2 obteve melhor resultado, com valores menores de desvio padrão e variância para o conjunto A e a arquitetura 1 apresentou melhor resultado para o conjunto B.

**Palavras-chave:** *Redes Neurais, Capitalização, Indicadores Econômicos.*

<sup>1</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [richard.hernani@gmail.com](mailto:richard.hernani@gmail.com)

<sup>2</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [leandro.ferreira@unifal-mg.edu.br](mailto:leandro.ferreira@unifal-mg.edu.br)

<sup>3</sup>ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha – MG. Email: [marcal.unifal@yahoo.com.br](mailto:marcal.unifal@yahoo.com.br)

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

# Estudo de raiz unitário de área colhida de alguns cultivos de 1931 a 2010

Valiana Alves Teodoro <sup>1 4</sup>

Djair Durand Frade Ramalho <sup>2 4</sup>

Taciana Villela Savian <sup>3 4</sup>

Na área colhida de laranja, milho, feijão, café, banana e arroz do Brasil houve um grande aumento nos últimos anos, por apresentar valores crescentes ao longo dos anos de 1931 a 2010(IPEA). Fazendo o teste da raiz unitária, para as séries, com o modelo AR(1) sem constante, todos os valores  $\tau$  foram menores que o  $\tau$  críticos a 1%, 5% e 10%, somente a série de milho que tem raiz unitária a 1%. Para o modelo AR(1) com constante, todos os valores  $\tau$  foram menores que os  $\tau$  críticos. Para o modelo AR(1) com tendência, todos os valores  $\tau$  foram menores que o  $\tau$  críticos, somente a série de café que tem raiz unitária a 1%. Fazendo o teste da raiz unitária, para a série com uma diferença, com o modelo AR(1) sem constante, todos os valores  $\tau$  foram maiores que o  $\tau$  críticos a 1%, 5% e 10%, somente a série de laranja que não tem raiz unitária a 1%. Para o modelo AR(1) com constante e o modelo AR(1) com tendência, todos os valores  $\tau$  foram maiores que o  $\tau$  críticos. Verificou que as séries de área plantada de laranja, milho, feijão, café, banana e arroz do Brasil de 1931 a 2010 são séries que não são estacionárias conferidas pelo teste da raiz unitária, assim esperado pois ao longo dos anos houve um crescimento de área plantada destes cultivos. E quando tomado uma diferença destas séries todas passam a ser estacionárias.

**Palavras-chave:** *Área colhida, Teste da raiz unitária, Série não - estacionária área colhida.*

---

<sup>1</sup>LCE-ESALQ/USP. Email: [valiana@usp.br](mailto:valiana@usp.br); [valianaalves@yahoo.com.br](mailto:valianaalves@yahoo.com.br)

<sup>2</sup>LCE-ESALQ/USP. Email: [djairdurand@gmail.com](mailto:djairdurand@gmail.com)

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

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

# COMPARAÇÃO ENTRE O ÍNDICE NACIONAL DE PREÇOS AO CONSUMIDOR AMPLO (IPCA) E A APLICAÇÃO DO ÍNDICE DE SAUERBECK NO SETOR DE HIGIENE PESSOAL, PERFUMARIA E COSMÉTICOS (HPPC).

Edna Maria F. P. Alves<sup>1 3</sup>

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

O Brasil é o terceiro mercado mundial de produtos de higiene pessoal, perfumaria e cosméticos, à frente de gigantes como a China e atrás apenas de Estados Unidos e Japão. Baseando-se na importância deste fato, foi utilizada uma amostra aleatória de nove produtos de uma grande empresa brasileira que comercializa produtos de HPPC para aplicar a metodologia de números índices e comparar com o Índice Nacional de Preços ao Consumidor Amplo (IPCA) do mesmo período analisado. Neste trabalho são utilizados cálculos de números índices compostos, agregados simples, pois são analisados vários produtos dentro de uma mesma classe, no caso HPPC, não sendo, portanto, necessário atribuir pesos diferentes para os cálculos. O índice utilizado foi proposto por Sauerbeck trabalha com a média aritmética dos relativos e é definido como  $IPS_{0,t} = \sum_{i=1}^n p_{0,t}^i / n$ . Este trabalho obteve como resultado um índice muito acima do IPCA do período em estudo. Logo é possível concluir que apesar do setor de HPPC ter pouca influência no IPCA, o índice de Sauerbeck para o setor apresenta-se elevado quando comparado ao IPCA, impactando diretamente no orçamento do brasileiro.

**Palavras-chave:** Números índices, HPPC, IPCA.

---

<sup>1</sup>DEEST – UFOP: Universidade Federal de Ouro Preto, Ouro Preto-MG. Email: [ednairam@hotmail.com](mailto:ednairam@hotmail.com)

<sup>2</sup>DEEST – UFOP: Universidade Federal de Ouro Preto, Ouro Preto-MG. Email: [felipe\\_pontenova@yahoo.com.br](mailto:felipe_pontenova@yahoo.com.br)

<sup>3</sup>Agradecimento ao Profº. Dr. Fernando Luiz P. de Oliveira pela orientação.

# **Análise Conjunta aplicada a problemas de marketing**

**Júlia Floresta Gonçalves<sup>1</sup>**

**Edmilson Rodrigues Pinto<sup>2</sup>**

Análise Conjunta (AC) é uma técnica multivariada que consiste em avaliar como os consumidores desenvolvem preferências por produtos e serviços, ou seja, como reagem à determinada combinação de fatores. Dentre as aplicações da AC estão: prever a venda ou utilização de um novo produto, selecionar características a serem oferecidas e estabelecer preços. Deve ser feita a definição do problema de marketing a partir da qual será possível identificar quais serão as informações necessárias para conduzir o estudo de forma a obter os melhores resultados para a tomada de decisão. O objetivo do trabalho é apresentar a Análise Conjunta, aplicada em um projeto de pesquisa de marketing, e as ferramentas utilizadas para sua aplicação. A metodologia se restringe à planejamento de experimento e análise de regressão. O Planejamento de Experimentos é empregado para definir os dados que devem ser coletados, buscando maior precisão estatística na resposta e menor custo. A Análise de Regressão é utilizada para investigar a relação entre duas ou mais variáveis. Serão considerados apenas modelos de Regressão Linear Múltipla, onde  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$ , em que  $\beta_0$  é a interseção,  $\beta_j$ ,  $j=1,2,\dots,k$  são coeficientes de regressão,  $X_j$ ,  $j=1,2,\dots,k$  são as variáveis independentes e  $Y$  é a variável independente ou resposta. A AC é muito aplicada em pesquisas de marketing, pois além de estudar a preferência por produtos e serviços, pode servir para segmentar o mercado com base na semelhança de preferências.

---

<sup>1</sup> FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia –MG Email: julia.floresta@hotmail.com

<sup>2</sup> FAMAT – UFU: Universidade Federal de Uberlândia, Uberlândia –MG Email: edmilson@famat.ufu.br (Orientador)

# Modelo Hierárquico aplicado ao estudo da relação entre o teor de umidade e o turno de fabricação de um tipo de sabonete

Bruna Cassia Rodrigues<sup>1</sup>

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

Mariana Vieira<sup>3</sup>

**Resumo:** O controle do fator umidade em cosméticos tais como sabonetes é importante, pois, altas umidades podem causar a decomposição dos componentes e degradação do produto durante o processo de estocagem. Segundo a Agência Nacional de Vigilância Sanitária, o percentual de umidade não deve exceder 10%. No entanto, fatores externos podem interferir de maneira negativa na umidade. Este trabalho teve como objetivo verificar a influência do turno e duração da jornada de trabalho e estação do ano na produção de um sabonete com relação ao teor de umidade. Avaliou-se um conjunto de dados proveniente de 4619 amostras de sabonete de uma marca comercial, coletados em três turnos de uma empresa durante o período de julho de 2012 a abril de 2013. Cada turno era composto de uma jornada de 8 horas de trabalho iniciando-se o 1º turno às 6 horas. Avaliou-se também o tempo de trabalho decorrido, considerando possível cansaço de funcionário dentro de cada turno. Assim, utilizou-se para análise um modelo hierárquico considerando os fatores meses de produção, turno e tempo decorrido. A análise foi realizada no software R, no qual os fatores turno e meses de produção foram significativos ( $p < 0,05$ ). Pelo teste Tukey, o teor de umidade dos meses difere de acordo com a mudança da estação climática, sendo que nos meses de verão a umidade aumenta consideravelmente, sendo superior ao limite estabelecido. Quanto ao turno de trabalho, concluiu-se que o terceiro turno possui o menor teor de umidade. Assim, pode-se dizer que os sabonetes de maior qualidade são aqueles produzidos nos meses de estação de seca durante o 3º turno.

**Palavras-chave:** *Hierárquico, Tukey, jornada de trabalho, cosmético.*

---

<sup>1</sup> Graduanda em Engenharia de Produção-UFSJ. e-mail: brunarodrigues557@gmail.com

<sup>2</sup> Departamento de Zootecnia-UFSJ. e-mail: carlabrighenti@ufs.edu.br

<sup>3</sup> PIBIC/Jr-FAPEMIG. e-mail: marianav152013@gmail.com

# Análise da série mensal da precipitação pluviométrica em Lavras-MG.

Érica Fernanda da Cruz<sup>1 5</sup>

Tamara Aparecida Nogueira dos Anjos<sup>2 5</sup>

Thelma Sáfadi<sup>3 5</sup>

Adriano Carvalho Costa<sup>4 5</sup>

**Resumo:** O índice de precipitação pluviométrica interfere diretamente em atividades econômicas, clima, produção agrícola e abastecimento populacional, conhecer seu comportamento histórico assume um papel de grande importância quando se considera a questão do planejamento e do uso dos recursos hídricos. Neste contexto, objetivou-se neste trabalho analisar e propor um modelo à série do índice mensal de precipitação pluviométrica da cidade de Lavras-MG, no período de janeiro de 2000 à julho de 2014 utilizando as técnicas de Séries Temporais. Observou-se que a precipitação pluviométrica da cidade de Lavras apresenta uma sazonalidade anual, ajustando-se melhor e de uma forma parcimoniosa ao modelo SARIMA(0, 0, 0)(1, 0, 1)<sub>12</sub>.

**Palavras-chave:** *Séries Temporais, Chuva, Sazonalidade.*

---

<sup>1</sup>DEX – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: erica.89@bol.com.br

<sup>2</sup>DEX – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: tamarananjos@gmail.com

<sup>3</sup>DEX – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: safadi@dex.ufla.br

<sup>4</sup>DZO – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: acarvalhocosta@gmail.com

<sup>5</sup>Agradecimento à FAPEMIG e à CAPES pelo apoio financeiro.

# Modelo de Regressão Poisson com Processos Auto Regressivos Latentes na Média Para Número de Óbitos Devido a Doenças Respiratórias na Cidade de São Paulo

Thaís Claudino Viana<sup>1 3</sup>

Glaura C Franco<sup>2 3</sup>

Desde a Revolução Industrial a poluição atmosférica é um fator presente na vida do planeta: a queima de carvão mineral iniciou o ciclo de técnicas industriais poluentes, responsáveis pela degradação do meio ambiente e geradora das doenças respiratórias, que acometem milhares de pessoas em todo o país. A cidade de São Paulo é a capital brasileira que mais sofre com a poluição do ar e, visando este fato, decidiu-se estudar a relação desta com o número de óbitos por doenças respiratórias. Foram coletadas informações a respeito de componentes presentes na atmosfera e a temperatura, bem como o número de mortes por estas doenças durante 329 dias do ano de 1997, gerando uma série  $Z_t$ . Para a análise dos dados foram utilizadas técnicas estatísticas de previsão baseados no modelo autorregressivo de médias móveis linear generalizado (GLARMA), proposto por Davis et al. (2003). Este procedimento é uma extensão do modelo linear generalizado, que incorpora na média do processo, além das variáveis explicativas, a autocorrelação presente na variável resposta, que neste caso é uma série de contagem do número de óbitos por doenças respiratórias. Com o auxílio do software R e utilizando o nível de 5% de significância para os testes, verificou-se que o modelo para descrever esta série pode ser explicado por um modelo GLARMA Poisson com a variável temperatura, nível de CO do ar e um processo auto regressivo de ordem 3. A análise de resíduos mostrou que o modelo está bem adequado aos dados.

**Palavras-chave:** *GLARMA, poluição, autocorrelação, séries de contagem.*

---

<sup>1</sup>ICEX – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Aluna de graduação em Estatística. Email: [thais\\_tcv@hotmail.com](mailto:thais_tcv@hotmail.com)

<sup>2</sup>ICEX – UFMG: Universidade Federal de Minas Gerais, Belo Horizonte-MG. Mestre e Doutora em Estatística e Orientadora deste trabalho. Email: [glaura@est.ufmg.br](mailto:glaura@est.ufmg.br)

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

# ANÁLISE DESCRITIVA ESPACIAL DO IDHM 2010, GINI 2010 E IDEB 5º ANO DE 2011 DO ESTADO DE MINAS GERAIS

Karolyna Ferreira Pires<sup>1</sup>

Ednaldo Carvalho Guimarães<sup>2</sup>

Os índices e coeficientes analisados são de suma importância para a sociedade e para os planejamentos das políticas públicas. O objetivo deste trabalho é fazer a análise descritiva espacial dos dados do IDHM de 2010, do GINI de 2010 e do IDEB do 5º ano de 2011, a partir de uma amostra de 300 cidades do estado de Minas Gerais. As análises descritivas espaciais foram feitas por meio de resultados obtidos de softwares estatísticos. Na análise foram verificadas médias, desvios padrões, coeficientes de assimetria ( $C_s$ ), coeficientes de curtose ( $C_k$ ) e gráficos de histograma e posição. A média do IDHM foi de 0,68087, com desvio padrão de 0,056228,  $C_s = -0,230$  e  $C_k = -0,404$ . O GINI teve média de 0,496765, desvio padrão de 0,0494763,  $C_s = 0,082$  e  $C_k = 0,785$ . E o IDEB obteve média de 5,769, desvio padrão de 0,5666,  $C_s = -0,104$  e  $C_k = -0,501$ . Os dados apresentam tendência à distribuição normal. Conclui-se pelas análises feitas que Minas Gerais está regular em relação aos índices IDHM e GINI e o IDEB está próximo de alcançar a meta de 2022. Esse tipo de análise é importante para posterior análise da dependência espacial.

**Palavras-chave:** *Análise exploratória, Estatística espacial, Índice de desenvolvimento.*

---

<sup>1</sup>UFU: Universidade Federal de Uberlândia, Uberlândia-MG. Email: *karolynaferreira@hotmail.com*. Bolsista PIBIC/FAPEMIG.

<sup>2</sup>UFU: Universidade Federal de Uberlândia, Uberlândia-MG. Email: *ecg@ufu.br*. Professor orientador.

# Comparação de modelos: formação de critérios e alternativas

Elayne Penha Veiga<sup>1 2</sup>

A modelagem matemática e estatística é praticamente intrínseca a qualquer fenômeno observado cientificamente. Diante de uma complexidade da natureza dos dados observados em quaisquer destes fenômenos vários modelos podem ser formulados, havendo então a necessidade de compará-los para decidir qual melhor representa estes dados. A construção de um critério de informação, comumente utilizado para comparação de modelos estatísticos baseia-se nos conceitos de Informação e Entropia. Eles quantificam a informação perdida no ajuste do modelo, penalizados por diferentes quantidades. Uma destas informações é a Informação de Kullback-Leibler, que é fundamental na construção das medidas de comparação de modelos mais comuns tais como AIC (Akaike, 1974) e BIC (Schwarz, 1978). Bem como uma opção baseada no preditivismo: a Medida L (Ibrahim e Laud, 1994), baseada em conceitos preditivos como a Densidade Preditiva em sua definição, e com este, comparando o que é predito ao que  $\tilde{A}$  observado para fazer a escolha entre modelos.

**Palavras-chave:** *Informação de Kullback-Leibler, Entropia, Comparação de modelos.*

---

<sup>1</sup>UFLA: Universidade Federal de Lavras, Lavras-MG. Email: elayne@posgrad.ufla.br

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

# Utilização da técnica de Bootstrap na análise das variáveis determinantes no cálculo da permeabilidade do filtro de polipropileno

Rosilanny Soares Carvalho, e-mail:rosilannysoares@gmail.com<sup>1</sup>

João Carlos Gonçalves, e-mail:joao.carlos@ifnmg.edu.br<sup>2</sup>

Maria de Fátima Ferreira Almeida, e-mail:fairam2006@yahoo.com.br<sup>3</sup>

**Resumo:** A técnica de reamostragem *Bootstrap*, proposta por Efron (1979) é uma boa alternativa no processo de inferência estatística para o parâmetro populacional. Esta técnica é amplamente utilizada em testes de hipóteses estatísticos para cálculo da incerteza a cerca dos parâmetros populacionais, por permitir que com uma quantidade reduzida de informações coletadas em laboratório possa gerar várias réplicas independentes para melhor estimação do verdadeiro parâmetro. Neste método supõe-se que a distribuição que originou os dados para as  $k$  amostras bootstrap, foram geradas utilizando esse modelo a partir dos parâmetros estimados com os dados da amostra original. Neste trabalho foi realizado um experimento para determinar a permeabilidade de um filtro de polipropileno que compõe a parte cônica de um ciclone com mangas para determinar a permeabilidade. A partir da aplicação da técnica *bootstrap*, implementada no software R, foi feito simulações das variáveis: queda de pressão no início e no fim da parte filtrante, com o intuito de gerar réplicas dos dados experimentais obtidos em laboratório e estimar a média populacional. A vantagem do método bootstrap é permitir a redução do tempo gasto para obter os valores das variáveis, além de permitir extrapolações.

**Palavras-chave:** *Bootstrap, Reamostragem, Permeabilidade.*

[1] EFRON, B. ; TIBSHIRANI, R. . *An Introduction to the Bootstrap*. London: Chapman & Hall, 1993.

[2] RODRIGUES, M. V. *Análise do desempenho de um ciclone com mangas: efeito da permeabilidade do meio filtrante*. Uberlândia, 2001. Dissertação (Mestrado em Engenharia Química). Universidade Federal de Uberlândia: Programa de Pós-Graduação em Engenharia Química.

# **Análise estatística e relação entre dados referentes a emissão de gases poluidores da camada de ozônio e taxas de crescimento das exportações**

**Gabriel Silveira Viana de Carvalho**<sup>1 3</sup>

**Keline Moraes Balieiro**<sup>2 3</sup>

**Maria de Fátima Ferreira Almeida**<sup>3 3</sup>

Desde o início do processo de industrialização se tem buscado novos processos produtivos visando maior quantidade e qualidade de produtos e serviços. Assim durante muito tempo não houve preocupação com o que isso impactaria no meio ambiente. Um dos principais males causados pela indústria é a emissão dos gases Clorofluorcarbonos (CFCs). Tendo em vista este cenário, foram buscados dados para avaliar se existe influência das taxas de crescimento exportações de produtos e serviços no aumento da emissão dos CFCs, utilizando o banco de dados de séries estatísticas do IBGE de 1996 a 2009.

A partir destes dados foi realizada a análise de covariância e correlação para verificar uma relação entre as variáveis: taxas de crescimento das exportações de produtos e serviços e aumento da emissão dos gases CFCs. Em seguida, observando que havia uma correlação, foi realizada uma análise de regressão que de acordo com o comportamento do gráfico de dispersão, ajustou uma regressão polinomial quadrática entre os dados. Observou que existe um ponto máximo de emissão de CFCs em função do aumento das taxas de exportações de produtos e serviços. Quando a taxa de exportação provoca o maior aumento na emissão de CFCs, interpreta-se que para aumento das exportações existe um aumento da emissão de gases CFCs até um valor limite dessas taxas a partir das quais o aumento não interfere mais no aumento da emissão de CFCs, porque a partir daí começa-se o ciclo de produção mais consciente. Isso se deve, possivelmente, ao controle ambiental feito por meio de aplicação das leis ambientais que mantêm ou buscam reduzir os índices de poluição, ou devido a outras variáveis que não foram levadas em conta na análise e podem ter influenciado. O estudo dessas variáveis poderão servir de base para pesquisas futuras.

**Palavras-chave:** CFC, exportações, regressão .

---

<sup>1</sup>IFNMG : Instituto federal do Norte de Minas Gerais, Montes Claros-MG. Email: gabrielviana77@yahoo.com.br.

<sup>2</sup>IFNMG : Instituto federal do Norte de Minas Gerais, Montes Claros-MG. Email: balieirokeline0@gmail.com.

<sup>3</sup>IFNMG : Instituto federal do Norte de Minas Gerais, Montes Claros-MG. Email: fairam2006@yahoo.com.br.

# **Determinantes das arrecadações do seguro de responsabilidade civil no Brasil: uma análise econométrica no período de 2004 a 2011.**

**Pablo Cescon Portes<sup>1 3</sup>**

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

**Resumo:** O seguro de responsabilidade civil tem como objetivo transferir as indenizações, relacionadas a responsabilidade civil, impostas judicialmente ao segurado para a seguradora. Com o crescimento do mercado desse tipo de seguro no Brasil, o presente trabalho tem por objetivo apresentar um modelo econométrico com variáveis independentes que possam ter relações com esse crescimento, são elas: número de processos judiciais por ano no Brasil, gasto das seguradoras com sinistro relacionados a responsabilidade civil e o PIB per capita. A relação entre as variáveis foi estudada por meio do uso de ferramentas estatísticas, mais precisamente, análise de regressões lineares simples e múltiplas. O modelo ajustado, para o PIB per capita, apresentou os melhores ajustes, um coeficiente de determinação ( $R^2$ ) de 0,8746, tal estatística denota que a variável explica a variável dependente em 87,46% dos casos, além disso, a estatística apresentou uma significância estatística inferior a 5%, confirmando a confiabilidade do ajuste. A regressão linear múltipla não apresentou boa qualidade nos ajustes e os sinais das variáveis independentes não ficaram de acordo com o esperado. Com isso, constatou-se que o PIB per capita foi a variável que mais se relacionou com o aumento do total de prêmio pago anualmente para as seguradoras com o seguro de responsabilidade civil. As outras variáveis não tiveram, estatisticamente e probabilisticamente, resultados significativos.

**Palavras-chave:** Seguro de responsabilidade civil, Influência seguro de responsabilidade civil.

<sup>1</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha-MG. Email: *pablo.portes@hotmail.com*

<sup>2</sup> ICSA – UNIFAL: Universidade Federal de Alfenas, Varginha-MG. Email: *gabriel.pessanha@unifal-mg.edu.br*

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

# Comparação de desempenho acadêmico entre cotista e não cotista na Universidade Federal de Uberlândia

João Flávio A. Silva<sup>1,5</sup>  
Thacyo E. De Villa<sup>2,5</sup>  
Janser Moura Pereira<sup>3,5</sup>  
Quintiliano S. S. Nomelini<sup>4,5</sup>

O presente estudo teve como objetivo comparar o desempenho entre os alunos ingressantes, na Universidade Federal de Uberlândia, por meio do PAAES e os que ingressaram por intermédio do vestibular. Considerou-se como desempenho as notas nas disciplinas no decorrer dos anos de 2011 e 2012, dos alunos ingressantes no ano de 2011. Foi utilizado o teste de Mann-Whitney para averiguar se existe diferença significativa das notas em relação à forma de ingresso. Ao nível de significância de 5%, foi possível observar que não existe diferença significativa dos desempenhos entre ingressantes do PAAES e do Vestibular para as áreas do conhecimento “Ciências Médicas e da Saúde”, “Ciências Humanas” e para um grupo contendo todas as áreas, já para as áreas “Ciências Matemáticas e Naturais”, “Engenharias e Computação”, “Ciências Biológicas”, “Ciências Agrônômicas e Veterinárias”, “Ciências Socialmente Aplicáveis” e “Linguagens e Artes” foram detectadas diferenças significativas.

**Palavras-chave:** *Ingressante, Inferência não-paramétrica, Rendimento acadêmico.*

---

1 FAMAT - Universidade Federal de Uberlândia. Email: joao.1988@hotmail.com

2 FAMAT - Universidade Federal de Uberlândia. Email: thacyo@est.ufu.br

3 FAMAT - Universidade Federal de Uberlândia. Email: janser@famat.ufu.br

4 FAMAT - Universidade Federal de Uberlândia. Email: quintiliano@famat.ufu.br

5 Agradecimento à FAPEMIG e a CAPES pelo apoio financeiro, e a DIRAC pela concessão dos dados.

# Análise da influência do café no peso de animais (ratos) por meio da análise de variância univariada

Wederson Leandro Ferreira<sup>1 5</sup>

Naje Clécio Nunes da Silva<sup>2 5</sup>

Marcelo Ângelo Cirillo<sup>3 5</sup>

Juracy Mendes Moreira<sup>4 5</sup>

A análise de variância univariada é uma metodologia estatística que pode ser utilizada para analisar dados longitudinais, desde que a matriz de covariâncias dos erros seja esférica. Ao realizar a análise de variância de um experimento no esquema de parcela subdividida no tempo é necessário que a pressuposição de variâncias iguais e correlações nulas seja satisfeita, caso contrário, para que a distribuição  $F$  seja exata é necessário que haja correção dos graus de liberdade no tocante às fontes de variações da subparcela. O objetivo deste trabalho foi analisar a influência do uso do café no ganho de peso médio de animais (ratos) por meio da análise de variância corrigida (ou univariada) por duas formas distintas e a forma clássica de análise. Os dados são provenientes de um delineamento inteiramente casualizado, 10 repetições, sendo os tratamentos constituídos por 6 dietas alimentares, sendo 3 sem extrato de café e 3 com extrato. O teste de Mauchly indicou que a matriz de covariâncias do modelo não satisfaz ao critério de esfericidade, não possuindo variâncias iguais e correlações nulas, desta forma, foi corrigido os graus de liberdade, ficando constatado que haverá conclusões distintas da análise de variância tradicional e a corrigida por uma das formas. Pelo teste de Scott Knott ficou constatado que a medida que o tempo evolui a dieta NC (com café), tendeu a diminuir o peso médio dos animais, quando comparado com a dieta N, que não recebeu o café.

**Palavras-chave:** *Parcela subdividida no tempo, teste de esfericidade de Mauchly, Café.*

---

<sup>1</sup> PPEEA – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: wedelean@gmail.com

<sup>2</sup> PPEEA – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: naje.silva@ifac.edu.br

<sup>3</sup> PPEEA – UFLA: Universidade Federal de Lavras, Lavras-MG. Email: macufla@dex.ufla.br

<sup>4</sup> UniRV: Universidade de Rio Verde, Rio Verde-GO, Email: juracimendesmoreira@yahoo.com

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

# Controle Estatístico Não Paramétrico de Processos para Medidas Individuais: Comparação de Desempenho de Núcleos Estimadores e *Bootstrap*

Alan de Paiva Loures<sup>1</sup>  
Lupércio França Bessegato<sup>2</sup>  
Fernando Luis Pererira de Oliveira<sup>3</sup>

Em geral, espera-se que um processo de produção seja estável ou replicável, ou seja, que ele tenha capacidade de operar com pequena variabilidade em torno de dimensões-alvo das características de qualidade do produto. O controle estatístico do processo (CEP) é uma poderosa coleção de ferramentas de resolução de problemas que são importantes para se alcançar a estabilidade do processo e a melhoria de sua capacidade, através da redução de variabilidade. A operação dos gráficos de controle consiste na coleta periódica de itens produzidos, analisando-os de acordo com alguma característica de interesse. O gráfico contém uma linha central (LC), representando o valor médio da característica de qualidade, e duas outras linhas horizontais, chamadas limite superior de controle (LSC) e limite inferior de controle (LIC). Escolhe-se a amplitude do intervalo entre LSC e LIC de maneira que, quando o processo estiver operando sob controle, praticamente todos os pontos amostrais estejam em seu interior.

Neste trabalho é considerada a situação não paramétrica para medidas individuais em que a função de distribuição subjacente,  $F$  é desconhecida. Assim discutimos metodologias para a escolha do parâmetro de suavização das estimativas dos limites de controle baseadas em núcleo estimador e bootstrap. São simuladas amostras de Fase I de algumas distribuições para determinar os limites de controle.

**Palavras-chave:** Núcleo estimador; Técnica bootstrap; Taxa de falso alarme.

---

<sup>1</sup> DE/ICE – UFJF. e-mail: alan.loures@ice.ufjf.br

<sup>2</sup> DE/ICE – UFJF. e-mail: lupercio.bessegato@ufjf.edu.br

<sup>3</sup> DEEST – UFOP. E-mail: fernandoluzest@gmail.com

<sup>4</sup> Os autores agradecem à FAPEMIG e a PROPESQ/UFJF ao apoio financeiro a esta pesquisa

## **Análise de parâmetros na usinagem de compósitos metálicos por meio de comparações múltiplas**

**Paulo Cesar de Resende Andrade**<sup>1</sup>  
**Milla Caroline Gomes**<sup>2</sup>  
**Ricardo Augusto Gonçalves**<sup>3</sup>

O presente trabalho tem como objetivo avaliar os efeitos da velocidade de corte e da pressão de aplicação do fluido de corte no desempenho do processo de usinagem com base nos experimentos realizados no trabalho Taguchi's Technique in Machining of Metal Matrix Composites (SHETTY; et. al., 2008) utilizando abordagem estatística.

Os parâmetros avaliados foram velocidade de corte e pressão de aplicação do fluido de corte. A rugosidade do material usinado, o desgaste da ferramenta de corte, e as componentes de avanço e de corte da força de usinagem consistem nos parâmetros de saída utilizados para analisar o desempenho do processo. O material da peça usinada foi um compósito de matriz metálica, constituído por uma matriz de liga de alumínio 6061 e partículas de carboneto de silício como reforço. A operação de usinagem selecionada foi o torneamento. A ferramenta de corte empregada foi o nitreto cúbico de boro. O fluido de corte empregado foi o vapor d'água saturado.

Neste trabalho foi adotada a combinação dos parâmetros velocidade de corte e pressão de vapor na determinação dos tratamentos planejados em um delineamento inteiramente casualizado. A partir dessa combinação os dados foram coletados para cada parâmetro de saída. Esses dados foram submetidos à análise de variância e a um teste bayesiano de comparações múltiplas e ao teste de Tukey, a um nível de significância de 5%, utilizando o software R. A velocidade de corte e a pressão de aplicação do fluido de corte foram variados em dois níveis, 45 e 101 m/min e 4 e 10 bar, respectivamente. Os resultados dos efeitos de cada tratamento foram analisados para cada um dos parâmetros de saída.

Um processo de usinagem com melhor desempenho apresenta menores valores do desgaste da ferramenta de corte, força de corte, força de avanço e rugosidade. De acordo com os resultados encontrados, similares em ambos os testes (bayesiano e Tukey), a menor rugosidade foi obtida a partir da aplicação da maior pressão de vapor e da maior velocidade de corte, que por sua vez, contribuíram para o maior desgaste da ferramenta de corte. Já as forças de avanço e de corte foram minimizadas quando utilizada a maior pressão de aplicação do fluido de corte. A partir deste trabalho foi verificado que tanto a velocidade de corte como a pressão de aplicação do fluido de corte possuem efeito sobre a usinabilidade do material.

**Palavras-chave:** comparações múltiplas, usinagem, compósitos metálicos.

1 ICT – UFVJM. e-mail: [paulo.andrade@ict.ufvjm.edu.br](mailto:paulo.andrade@ict.ufvjm.edu.br)

2 ICT – UFVJM. e-mail: [millagomes15@yahoo.com.br](mailto:millagomes15@yahoo.com.br)

3 ICT – UFVJM. e-mail: [ricardo.augusto@ict.ufvjm.edu.br](mailto:ricardo.augusto@ict.ufvjm.edu.br)

4 Agradecimento ao CNPq pelo apoio financeiro.

# Técnicas Estatísticas Aplicadas a Mineração de Dados - Classificação Bayesiana

Bárbara de Oliveira Aguiar <sup>1</sup>

Fernanda Ires Silva <sup>2</sup>

Daniel Antunes Rocha <sup>3</sup>

José Francisco C. de Jesus <sup>4</sup>

Luciana Balieiro Cosme <sup>5</sup>

Maria de Fátima Ferreira Almeida <sup>6</sup>

A computação dispõe da mineração de dados como uma prática útil em técnicas de recuperação de informação, inteligência artificial, reconhecimento de padrões e estatística, para procurar correlações entre diferentes dados. Durante o processo de mineração, diversas técnicas devem ser testadas e combinadas para que comparações possam ser feitas e então a melhor técnica ou combinação delas seja utilizada. Existem cinco técnicas gerais de mineração de dados, são elas: classificação, estimativa, previsão, análise de afinidades e análise de agrupamentos [Carvalho, 2005]. A proposta deste trabalho é analisar o método de classificação, dando ênfase apenas na Classificação Bayesiana, vale ressaltar que há outras formas de classificação. As técnicas de classificação podem ser supervisionadas e não-supervisionadas e são usadas para prever valores de variáveis do tipo categóricas. A Classificação Bayesiana é baseada no teorema de Thomas Bayes, que segundo Morettin, é uma das relações mais importantes sobre probabilidade. O teorema diz que é possível encontrar a probabilidade de certo evento ocorrer, dada a probabilidade de outro evento que já ocorreu. Comparativos mostram que os algoritmos Bayesianos, naive Bayes, obtiveram resultados compatíveis com os métodos de árvore de decisão e redes neurais. Por sua simplicidade e o alto poder preditivo, é um dos algoritmos mais utilizados, e segundo estudos apresentam precisão e agilidade quando aplicados a grande base de dados [ZHANG, 2004].

**Palavras-chave:** *Mineração de Dados, Classificação Bayesiana, Algoritmos.*

---

<sup>1</sup>Instituto Federal do Norte de Minas Gerais, *Campus* Montes Claros. Email: *barbara.oliveira.aguiar@gmail.com*

<sup>2</sup>Instituto Federal do Norte de Minas Gerais, *Campus* Montes Claros. Email: *iresnanda@gmail.com*

<sup>3</sup>Instituto Federal do Norte de Minas Gerais, *Campus* Montes Claros. Email: *danielantrocha@gmail.com*

<sup>4</sup>Instituto Federal do Norte de Minas Gerais, *Campus* Montes Claros. Email: *rj45francis@gmail.com*

<sup>5</sup>Instituto Federal do Norte de Minas Gerais, *Campus* Montes Claros. Email: *luciana.balieiro@ifnmg.edu.br*

<sup>6</sup>Instituto Federal do Norte de Minas Gerais, *Campus* Montes Claros. Email: *fatima.almeida@ifnmg.edu.br*

**Utilização da Krigagem Indicativa na geração de mapas para identificação das áreas onde o Indicador de Densidade de Ocupação econômica das Terras (DOET) da Região do Norte de Minas Gerais atingiu o nível satisfatório de acordo com a média do DOET das cidades analisadas.**

Maria de Fátima Ferreira Almeida<sup>1</sup>  
Gérson Rodrigues dos Santos<sup>2</sup>  
Magaly Stefânia Almeida<sup>3</sup>  
Wagner Rogério pinheiro<sup>4</sup>  
Alex da Silva Santos<sup>5</sup>

O estudo da agricultura no Estado e nas mesorregiões mineiras é realizado mensurando-se os efeitos área, rendimento, composição e localização geográfica ( BASTOS; GOMES, 2011). O uso de técnicas geoestatísticas para obter mapas dos índices de Densidade de Ocupação Econômica da terra e do Índice de Densidade de Ocupação da Terra por Cultura, permite localizar as subáreas e municípios onde há melhor ocupação da terra com agricultura e quais culturas são predominantes em cada subregião da área estudada. De acordo com os autores, Isaak e Srivastava (1989) e Landim (2003), por meio da Krigagem pode-se obter a previsão do valor pontual de uma variável regionalizada em um local específico dentro do espaço geométrico, através de um procedimento exato de interpolação. O indicador Densidade de Ocupação Econômica das Terras (DOET), leva em consideração as variáveis áreas cultivadas, áreas ocupadas por unidade de preservação e aquelas inundadas, permitindo acompanhamento indireto da produtividade e possibilitando observar prováveis riscos de contaminação dos solos ( SCOLFORO et al, 2008). Este trabalho apresenta os mapas de probabilidades feito por Krigagem Indicativa do DOET por cidade do Norte de Minas Gerais em conjunto com o DOTC para as três principais culturas predominantes na região, banana, milho e feijão, utilizando como ponto de corte a média do DOET e DOTC das cidades analisadas com o propósito de explicitar a organização espacial, a heterogeneidade existente em relação as principais culturas. Para isto, utilizou-se bancos de dados disponibilizados pelo (IBGE, 2010) para o cálculo do DOET e para o georreferenciamento dos 50 municípios selecionados via amostragem aleatória probabilística extratificada das microrregiões.

**Palavras-chave:** *DOET, Krigagem Indicativa, Mapeamento.*

---

<sup>1</sup> DET - UFV. e-mail:fatima.almeida@ifnmg.edu.br

<sup>2</sup> DET – UFV. e-mail:gerson.santos@ufv.br

DEQ-FACIT- e-mail: magalystefania@yahoo.com.br

<sup>3</sup> DET – UFV. e-mail:wagner2235@gmail.com

<sup>2</sup> DET – UFV. e-mail:alxcart@ibge.gov.br

# Seleção de um modelo de regressão binomial no estudo da predação do ácaro *Euseius concordis* (CHANT, 1959) na demanda de concorrente natural do cafeeiro

Gilberto Rodrigues Liska<sup>1 6</sup>

Erika Carla da Silveira<sup>2 6</sup>

Marcelo Ângelo Cirillo<sup>3 6</sup>

Guido Gustavo Humada-Gonzalez<sup>4 6</sup>

Estevan Rodrigues Liska<sup>5 6</sup>

O ácaro-vermelho *Oligonychus ilicis* (McGregor, 1917) tem se mostrado como um importante agente causador de problemas na cultura do cafeeiro (*Coffea spp.*). O controle químico tem sido amplamente utilizado no controle de *O. ilicis*, porém, a existência dos efeitos adversos de sua aplicação tem gerado preocupação quanto à sua utilização. Uma alternativa para esse método é o controle biológico das pragas por predadores naturais, entre eles os ácaros da espécie gênero *Euseius concordis* têm sido relatadas como predadores eficientes no controle biológico de várias espécies de ácaros fitófagos, em diversos cultivos. Nesse sentido, o presente trabalho objetivou estudar a proporção de ataque do ácaro predador *Euseius concordis* na demanda da presa *Oligonychus ilicis*, selecionando o modelo de regressão binomial mais apropriado para essa situação. Os resultados apontam que o modelo de regressão binomial logit apresentou os melhores valores dos critérios de adequabilidade AIC, BIC e Deviance, sendo, portanto, o modelo mais adequado para explicar a relação de predação/demanda do ácaro *E. concordis* no controle do *O. ilicis*. O desempenho de *E. concordis* em densidades superiores a 30 ácaros/3cm<sup>2</sup> é menor do que 50%. A partir da densidade 45 ácaros/3cm<sup>2</sup> a taxa de predação do ácaro *E. concordis* apresentou ligeiro decréscimo e tendência de constância.

**Palavras-chave:** Modelos lineares generalizados, *Coffea arabica*, controle biológico, Phytoseiidae, Tetranychidae .

---

<sup>1</sup>DEX - Universidade Federal de Lavras. Email: [gilbertoliska@hotmail.com](mailto:gilbertoliska@hotmail.com)

<sup>2</sup>DEX - Universidade Federal de Lavras. Email: [erika.silveira@yahoo.com.br](mailto:erika.silveira@yahoo.com.br)

<sup>3</sup>DEX - Universidade Federal de Lavras. Email: [marcelocirillo@gmail.com](mailto:marcelocirillo@gmail.com)

<sup>4</sup>DEX - Universidade Federal de Lavras. Email: [gustavohumad@hotmail.com](mailto:gustavohumad@hotmail.com)

<sup>5</sup>ICEX - Universidade Federal de Alfenas. Email: [estevanliska@gmail.com](mailto:estevanliska@gmail.com)

<sup>6</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

# Comparação da distribuição gama-normal com a odd-log logística-normal

Altemir da Silva Braga<sup>1 3</sup>

Edwin M. Marcos Ortega<sup>2 3</sup>

A distribuição normal é uma das mais destacada dentro da área da estatística, porém tal distribuição não se ajusta bem a dados que apresentam assimetria. Isto ocorre, porque a essa distribuição possui apenas os dois primeiros momentos. Por isso, muitos estudos são realizados com a finalidade de criar novas famílias de distribuições que acomodem, tanto a medida de assimetria, quanto a curtose dos dados. Consequentemente, tais famílias passam a ter mais parâmetros e exigir mais recursos computacionais. Pode-se encontrar na literatura algumas extensões da distribuição normal como, por exemplo, a skew-normal, a beta-normal, a kumaraswamy-normal, a gama-normal e a odd log-logística-g. Proposta por Ayman Alzaatreh, Felix Famoyeb e Carl Leea (2013) a família de distribuição de probabilidade gama-normal é um classe de distribuição paramétrica que apresenta uma extensão da distribuição normal por meio de um único parâmetro (parâmetro de assimetria). Enquanto que a família de distribuições odd log-logística-g foi proposta por Nilton et al. (2013) com a ideia de criar uma nova família de distribuição que pudesse estender qualquer distribuição contínua. Assim, pensou-se em utilizar a distribuição normal como função base na distribuição odd log-logística-g e ajustar as distribuições gama-normal e odd log-logística-normal utilizou-se o método da máxima verossimilhança por meio do pacote *optim* do software R. Dessa forma, conduziu-se este estudo, com o objetivo de ajustar tais distribuições e compará-las utilizando os critérios de comparação *AIC* e *BIC*. Portanto, após a modelagem estatística, a distribuição odd log-logística-normal ajustou-se melhor aos dados.

**Palavras-chave:** *Ajuste; Assimetria; Comparação; Gama; Chances.*

---

<sup>1</sup>USP/ESALQ: Doutorando da Escola Superior de Agricultura Luiz de Queiros, Piracicaba-SP. Email: *altemir@usp.br*

<sup>2</sup>USP/ESALQ: Professor Titular da Escola Superior de Agricultura Luiz de Queiros, Piracicaba-SP. Email: *edwin@usp.br*

<sup>3</sup>FAPAC: Agradecimento à Fundação de Amparo à Pesquisa do Estado do Acre pelo apoio financeiro.

## Efeito da Idade Relativa em Atletas Olímpicos de Lutas

Thayane da Silva Campideli<sup>1</sup>

Emerson Filipino Coelho<sup>1,2</sup>

Renato Melo Ferreira<sup>1,2</sup>

Francisco Zacaron Werneck<sup>1,2</sup>

No alto nível, observa-se maior representação de atletas nascidos nos primeiros meses do ano, devido aos efeitos da idade relativa (EIR) - possíveis vantagens que os atletas relativamente mais velhos levam em relação aos seus pares. Porém, o EIR carece de investigação nos esportes de combate. O presente estudo analisou o quartil de nascimento dos atletas olímpicos das modalidades de combate de ambos os sexos que participaram das Olimpíadas de Londres 2012. O mês de nascimento de 1163 atletas (boxe = 287; judô = 392; luta olímpica = 352; taekwondo = 132) foi categorizado em quartis (1°Q: Jan-Mar; 2°Q: Abr-Jun; 3°Q: Jul-Set; 4°Q: Out-Dez), a partir do site [www.london2012.com/athletes](http://www.london2012.com/athletes). Para a análise dos dados, empregou-se o teste Qui-Quadrado ( $X^2$ ) e nível de significância de 5%. Observou-se maior representação de atletas nascidos no 1°Q (27,2%) e 3°Q (27,1%) comparados ao 2°Q (22,6%) e 4°Q (23,1%) ( $X^2= 8,491$ ;  $df=3$ ;  $p=0,04$ ). Por outro lado, o EIR não foi confirmado quando a análise foi feita em cada modalidade separadamente: Boxe (26,8% vs. 21,3% vs. 27,9% vs. 24,0%;  $X^2= 3,049$ ;  $df=3$ ;  $p=0,38$ ), Judô (27,3% vs. 24,2% vs. 28,1% vs. 20,4%;  $X^2= 5,694$ ;  $df=3$ ;  $p=0,13$ ), Luta Olímpica (26,7% vs. 21,6% vs. 27,3% vs. 24,4%;  $X^2= 2,818$ ;  $df=3$ ;  $p=0,42$ ) e Taekwondo (28,8% vs. 23,5% vs. 22,0% vs. 25,8%;  $X^2= 1,394$ ;  $df=3$ ;  $p=0,71$ ). Conclui-se que o EIR não foi observado de forma consistente nos atletas das modalidades de lutas que participaram das Olimpíadas de Londres 2012.

**Palavras-chave:** Efeito da idade relativa; atletas; esportes de combate.

---

<sup>1</sup>CEDUFOP: Centro Desportivo da Universidade Federal de Ouro Preto, Ouro Preto-MG. Email: [f.zacaron@cedufop.ufop.br](mailto:f.zacaron@cedufop.ufop.br)

<sup>2</sup>LABESPE: Laboratório de Estudos e Pesquisas do Exercício e Esporte

Agradecimento à UFOP pelo apoio financeiro.

# Representação de dados funcionais como curvas suaves

Mateus Aguiar Florentino<sup>1</sup>

Ana Paula Tremura Galves<sup>2</sup>

A análise estatística de uma amostra de  $n$  curvas - formadas por dados que são registrados densamente ao longo do tempo - é chamada de análise de dados funcionais (ADF). Técnicas de suavização serão apresentadas e discutidas neste trabalho com o intuito de melhor representar a curva em questão. O objetivo principal é promover uma melhor análise de curvas formadas por dados registrados ao longo do tempo. Dessa forma, os erros da amostra são tratados e obtém-se uma curva mais fiel para análise posterior.

Seja  $x(t)$  um vetor de tamanho  $n$  com os valores da função  $x$  avaliados nos pontos de observação  $t_1, \dots, t_n$  e considere  $y_j$  os dados observados da amostra. Um suavizador linear estima o valor da função  $x(t)$  através de uma combinação linear de observações discretas, ou seja,  $\bar{x} = \sum_{j=1}^n S_j(t) y_j$ . O comportamento do suavizador em  $t$  é determinado pelos pesos  $S_j(t)$ .

Um método comum de suavização com funções base é o ajuste de mínimos quadrados, o qual consiste em representar a função como uma combinação linear de  $k$  funções base conhecidas  $\phi_k$ , ou seja,  $x(t) = \sum_{k=1}^K c_k \phi_k(t)$ , onde  $c_k$  representa os coeficientes da função.

Pode-se concluir que o grau de suavidade que será aplicado nos dados  $y_j$  é determinado pelo número de funções base  $k$ . Porém, é importante que as funções base apresentem características que se relacionem com aquelas encontradas nas funções que serão estimadas.

**Palavras chave:** suavização, dados funcionais, curvas.

---

<sup>1</sup>UFU: Universidade Federal de Uberlândia. Uberlândia, MG.  
Email: [maguiiar@hotmail.com](mailto:maguiiar@hotmail.com)

<sup>2</sup>UFU: Universidade Federal de Uberlândia. Uberlândia, MG.  
Orientadora. Email: [anapaula@famat.ufu.br](mailto:anapaula@famat.ufu.br)

# Modelos Hierárquicos e Séries Temporais para estudo de Tráfego Rodoviário

Diogo Maia de Figueiredo<sup>1 2</sup>

É de extrema importância realizar estudos de tráfego em uma estrada ou rodovia, seja para saber as características que ela deve possuir, como barreiras, largura, acostamento e até mesmo equipamentos de sinalização e segurança; ou até mesmo para se ter uma estimativa do volume de veículos que percorrem todo dia, uma informação que é útil para Concessionárias que pretendem conseguir licitação e instalar postos de pedágio no local. Assim, dados reais foram coletados na MG050, principal via que liga MG e SP, durante 2011, 2012 e 2013 diariamente e servirão de base para um projeto de Extensão. Com a orientação do professor Doutor Gregório Saravia Atuncar, foi pensado em duas formas de se estudar o banco de dados. A primeira foi considerando premissa de independência entre os dias e com isso foi desenvolvido um Modelo Hierárquico, onde o Ano da Coleta, Mês e Dia da Semana foram os fatores do planejamento. O segundo modelo já estuda a correlação entre os dados e por isso foi considerado a interação entre eles, sendo necessário o uso de Séries Temporais. Após o ajuste dos dois modelos, uma comparação foi feita onde saiu as conclusões e aplicações de cada um dos ajustes.

**Palavras-chave:** *Séries Temporais, Modelos Hierárquicos, Tráfego, Trânsito.*

---

<sup>1</sup> Aluno graduando no oitavo Período de Estatística pela *Universidade Federal de Minas Gerais* (UFMG). E-mail: [diogo.mfigueiredo@hotmail.com](mailto:diogo.mfigueiredo@hotmail.com).

<sup>2</sup> Agradecimento ao professor Doutor Gregório Saravia Atuncar pela orientação do projeto.