

**MARIANE ALVES GOMES DA SILVA**

**MODELOS MISTOS NA SELEÇÃO ENTRE E DENTRO DE FAMÍLIAS  
DE CANA DE AÇÚCAR SOB O ENFOQUE BAYESIANO**

**Dissertação apresentada à  
Universidade Federal de Viçosa,  
como parte das exigências do  
Programa de Pós-Graduação em  
Estatística Aplicada e Biometria,  
para obtenção do título de *Magister  
Scientiae*.**

**VIÇOSA  
MINAS GERAIS - BRASIL  
2012**

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

S586m  
2012

Silva, Mariane Alves Gomes da, 1986-  
Modelos mistos na seleção entre e dentro de famílias de  
cana de açúcar sob o enfoque bayesiano / Mariane Alves  
Gomes da Silva. – Viçosa, MG, 2012.  
x, 59f. : il. ; 29cm.

Orientador: Luiz Alexandre Peternelli.  
Dissertação (mestrado) - Universidade Federal de Viçosa.  
Inclui bibliografia.

1. Cana-de-açúcar - Melhoramento genético. 2. Teoria  
bayesiana de decisão estatística. 3. BLUPIS. I. Universidade  
Federal de Viçosa. II. Título.


CDD 22. ed. 633.612

MARIANE ALVES GOMES DA SILVA

**MODELOS MISTOS NA SELEÇÃO ENTRE E DENTRO DE FAMÍLIAS  
DE CANA DE AÇÚCAR SOB O ENFOQUE BAYESIANO**


Dissertação apresentada à  
Universidade Federal de Viçosa,  
como parte das exigências do  
Programa de Pós-Graduação em  
Estatística Aplicada e Biometria,  
para obtenção do título de *Magister  
Scientiae*.

APROVADA: 16 de fevereiro de 2012.

  
Moysés Nascimento

  
Felipe Lopes da Silva

  
Sebastião Martins Filho

  
Luiz Alexandre Peternelli  
(Orientador)

***“A dúvida é o princípio da sabedoria”.***

***(Aristóteles)***

## **AGRADECIMENTOS**

À Deus por sempre se manifestar na minha vida quando mais preciso.

À minha família, pelo amor, incentivo e apoio incondicional, que nunca me deixou desistir dos meus sonhos.

À todos os amigos que conquistei no campus da UFV, em especial aos que pude estudar junto e me ensinaram que estudar “brincando” é a única maneira de levar o aprendizado para a vida toda.

À todos os professores que lecionaram neste curso e em especial ao professor Luiz Alexandre Peternelli, pela perfeita orientação, apoio e confiança.

À Universidade Federal de Viçosa e a Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (CAPES-REUNI) pela oportunidade de realização deste curso e bolsa de estudo concedida.

## **BIOGRAFIA**

MARIANE ALVES GOMES DA SILVA, filha de Luiz Sérgio Valente Gomes da Silva e Maria Izabel Alves Gomes, nasceu em 29 de outubro de 1986 em Alegre, Espírito Santo.

Em 2009, graduou-se no curso Estatística na Universidade Federal do Espírito Santo.

Em 2010, iniciou o curso de Mestrado em Estatística Aplicada e Biometria na Universidade Federal de Viçosa tendo defendido sua dissertação em 16 de fevereiro 2012.

## SUMÁRIO

LISTA DE TABELAS .....	vii
RESUMO .....	ix
ABSTRACT .....	x
1 INTRODUÇÃO.....	1
1.2 Objetivos .....	3
1.2.1 Objetivo Geral .....	3
1.2.2 Objetivos Específicos .....	3
2 REFERENCIAL TEÓRICO.....	4
2.1 Modelos Mistos .....	4
2.1.1 Enfoque Clássico .....	6
2.1.2 Enfoque Bayesiano .....	7
2.2 Método de Monte Carlo .....	9
2.3 Diagnóstico de Convergência.....	11
2.3.1 Critério de Raftery e Lewis (1992).....	12
2.3.2 Critério de Geweke (1992).....	13
2.4 BLUPIS .....	14
2.5 Informação de Parentesco .....	15
3 REFERÊNCIAS .....	16
CAPÍTULO I - ANÁLISE DE DADOS REAIS.....	19
1 INTRODUÇÃO.....	19
2 MATERIAL E MÉTODOS.....	20
2.1 Detalhes experimentais.....	20
2.2 Procedimentos utilizados.....	21
2.2.1 Modelos Mistos .....	21
2.2.2 Método REML/BLUP .....	22
2.2.3 Método Bayesiano .....	24
2.2.4 Comparação entre os enfoques clássico e bayesiano na obtenção dos efeitos genéticos .....	27
2.2.5 Informação de Parentesco .....	27
3 RESULTADOS E DISCUSSÕES.....	27
4 CONCLUSÕES.....	39
5 REFERÊNCIAS .....	39
CAPÍTULO II – DESENVOLVIMENTO DAS FUNÇÕES NO R .....	41

1 INTRODUÇÃO.....	41
2 MATERIAL E MÉTODOS.....	42
2.1 Procedimentos para a realização do algoritmo.....	43
3 RESULTADOS.....	55
4 CONCLUSÃO.....	58
5 REFERÊNCIAS.....	58

## LISTA DE TABELAS

- Tabela 1 - Esquema da análise de variância para o delineamento em blocos casualizados completos mostrando fonte de variação (FV), graus de liberdade (GL) e esperanças de quadrado médio (E(QM)).....**24**
- Tabela 2 - Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (sem informação de parentesco e priori não informativa) para a variável tonelada de colmos por hectare (TCH).....**30**
- Tabela 3 - Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (sem informação de parentesco e priori informativa) para a variável tonelada de colmos por hectare (TCH)..... **31**
- Tabela 4 - Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (com informação de parentesco e priori não informativa) para a variável tonelada de colmos por hectare (TCH).....**32**
- Tabela 5 - Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (com informação de parentesco e priori informativa) para a variável tonelada de colmos por hectare (TCH).....**33**
- Tabela 6 - Amplitude dos intervalos das estimativas e predições de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (BLUPIS sem informação de parentesco e priori não informativa) para a variável tonelada de colmos por hectare (TCH) com número diferente de repetições.....**35**

Tabela 7 - Matriz de Parentesco com 22 famílias..... **36**

Tabela 8 - Vetores dos efeitos genotípicos ( $\hat{g}_j$ ), do número de indivíduos a serem selecionados ( $n_k$ ) pela metodologia BLUPIS e número de indivíduos selecionados pelo valor genotípico (BLUP) para a característica tonelada de colmos por hectare (TCH) considerando o número de indivíduos a serem selecionados na melhor família ( $n_j$ ) igual a 50.....**38**

## RESUMO

SILVA, Mariane Alves Gomes da, M.Sc., Universidade Federal de Viçosa, fevereiro de 2012. **Modelos mistos na seleção entre e dentro de famílias de cana de açúcar sob o enfoque bayesiano.** Orientador: Luiz Alexandre Peternelli. Coorientadores: Fabyano Fonseca e Silva e Márcio Henrique Pereira Barbosa.

A base do agronegócio de cana-de-açúcar é o melhoramento genético. Pode ser mostrado que a estratégia ótima de seleção da planta seria através da predição de valores genotípicos usando o BLUP (Best Linear Unbiased Predictor) individual (BLUPI). Este procedimento usaria, simultaneamente, a informação de família e de indivíduos para a seleção. No entanto esse método dificilmente é usado nos programas de melhoramento devido a problemas operacionais relacionados à obtenção dos dados ao nível de planta. Recentemente uma alternativa operacionalmente mais prática foi proposta, e é denominada BLUPIS (BLUP individual simulado). Nesse caso os dados são coletados ao nível de parcela. Com isso é possível selecionar as melhores famílias e, posteriormente, simular o número de indivíduos a serem selecionados dentro das melhores famílias. Este trabalho teve como objetivo desenvolver um algoritmo para análise do BLUPIS sob o enfoque bayesiano, com diferentes definições de distribuições *a priori* na sua modelagem, no software estatístico R, para possível disponibilização ao usuário e compará-la com o método clássico REML/BLUP. Os resultados mostraram que o método BLUPIS com enfoque bayesiano realizado através do algoritmo construído junto ao programa R foi eficiente. O algoritmo levou em consideração a incerteza existente sobre todos os parâmetros do modelo, como também possibilitou o uso de priori informativa. O método bayesiano se mostrou mais eficiente, isto é, com efeitos genotípicos maiores e variâncias e herdabilidade menores, quando se consideraram no modelo a informação de parentesco e a distribuição da priori informativa.

## ABSTRACT

SILVA, Mariane Alves Gomes da, M.Sc., Universidade Federal de Viçosa, February, 2012. **Mixed models in selection between and within families of cane sugar under the bayesian approach.** Adviser: Luiz Alexandre Peternelli. Co-Advisers: Fabyano Fonseca e Silva e Márcio Henrique Pereira Barbosa.

The base of agribusiness of sugar cane is the breeding. Can be shown that the great strategy selection of plant would be through the prediction of genotypic values using BLUP (Best Linear Unbiased Predictor) individual (BLUPI). This procedure would use both the information of family and individuals for selection. However this method is rarely used in breeding programs because of operational problems related to obtain data plant. Recently an alternative operationally more practical was proposal and is called BLUPIS (BLUP individual simulated). In this case the data are collected at the level of plot. With this is possible to select the best families, and subsequently, to simulate the number of individuals to be selected within the best families. This work has as aims to develop an algorithm for analysis BLUP under the Bayesian focus with different settings of priors in your modeling, in statistical software R, for possible available to the user and compare it with the Classic REML / BLUP. The results showed that the method with Bayesian BLUPIS performed by the algorithm built with the R program was effective. The algorithm took into account the uncertainties on all parameters, as also allowed the use of a priori information. The Bayesian method is more efficient when considering the relationship information model and the distribution of the prior information (major genotypic effects and variances and heritability smaller).

## 1 INTRODUÇÃO

A composição genética de diversas culturas de plantas é, na atualidade, resultado de todo um processo de melhoramento que elas foram sendo submetidas durante décadas. Os melhoristas são responsáveis pelo fenomenal progresso genético de um vasto número de espécies, entre elas se encontra a cana-de-açúcar.

A cana-de-açúcar (*Saccharum spp.*) é uma *Poacea* cultivada principalmente como matéria prima para a produção de açúcar e álcool e segundo Cesnik (2004) a cana originou-se primeiramente nas ilhas do Arquipélago da Polinésia e no Brasil se adaptou muito bem ao clima e ao solo nordestino no ano de 1530.

A base do agronegócio de cana-de-açúcar é o melhoramento genético e o cultivar é a tecnologia de maior importância para aumento de sua produtividade. Programas de melhoramento de plantas são realizados com a intenção de selecionar os cultivares que obtêm melhor lucro para os produtores durante muitos anos, sendo de extrema importância o desenvolvimento e uso de técnicas estatísticas apropriadas para otimizar o processo de seleção.

No entanto, a seleção individual de clones (seleção massal) aplicada nos estágios iniciais dos programas, apesar de ser usualmente praticado (MATSUOKA *et al.*, 2005) tem se mostrado ineficiente (KIMBENG, COX, 2003). Esses autores discutem que a seleção de famílias seguida da seleção individual de clones, apresenta maior ganho genético do que a simples seleção de clones ou de famílias. Além disso, a economia de recursos seria muito maior.

Resende (2002) mostra que a estratégia de seleção ótima em cana de açúcar seria através da predição de valores genotípicos usando o BLUP (*Best Linear Unbiased Predictor*) individual (BLUPI). Este procedimento usaria, simultaneamente, a informação de família e de indivíduos para a seleção. Ainda assim esse método dificilmente é usado nos programas de melhoramento devido a problemas operacionais relacionados à obtenção dos dados a nível de indivíduos.

Para contornar esta situação Resende e Barbosa (2006) propuseram a seleção via BLUP individual simulado (BLUPIS), que se baseia nos efeitos genotípicos das famílias avaliadas a campo. A vantagem desse método é que ele

não requer a avaliação individual de plantas no experimento. Além disso, este método indica o número de indivíduos a serem selecionado por família, o número total de clones para ganho com seleção, e o número de famílias envolvidas com os indivíduos.

As equações de modelos mistos são adequadas para a realização do método BLUP, pois apresentam tanto fatores de efeitos fixos como aleatórios, fornecendo uma estrutura mais apropriada para os dados.

Na análise de modelos lineares mistos pela abordagem clássica, rotineiramente a predição/estimação baseia-se no procedimento REML/BLUP (máxima verossimilhança restrita/ melhor predição linear não viciada). Segundo Oliveira (2008), a seleção de famílias por meio deste método pode ser uma estratégia importante para identificar famílias com elevados valores genotípicos, onde haveria maior probabilidade de seleção de clones potenciais. Entretanto, através deste procedimento, a distribuição e variância dos estimadores não são conhecidas, de forma que questões referentes à precisão das predições e estimações não podem ser respondidas com rigor.

A análise Bayesiana tem a vantagem de permitir a incorporação de informações conhecidas (*a priori*), caso existam, enriquecendo o processo de inferência. Segundo Resende et. al. (2001) a análise Bayesiana de modelos lineares mistos baseia-se no conhecimento da distribuição *a posteriori* dos parâmetros a serem estimados, fato que possibilita a construção de intervalos de credibilidade exatos para as estimativas das variáveis aleatórias, efeitos fixos e componentes de variância que tendem a conduzir também a uma maior aproximação entre ganhos genéticos preditos e realizados com seleção; o que torna essa abordagem mais confiável do que o procedimento da análise clássica.

A inferência bayesiana trata os parâmetros populacionais como variáveis aleatórias e um método útil de simulação é o de Monte Carlo (gera números aleatórios de uma distribuição de probabilidade) que tem a finalidade de obter de forma facilitada os intervalos de credibilidade (NOGUEIRA, 2003).

Os métodos de Monte Carlo via cadeias de Markov (MCMC - *Markov Chain Monte Carlo*) são uma alternativa na solução de problemas complexos. Um algoritmo muito usado é o amostrador de Gibbs, que pode ser entendido como

uma implementação prática, a partir do conhecimento das distribuições condicionais, na determinação da distribuição conjunta a posteriori, se ela existir.

Segundo Resende (2001) as estimativas dos parâmetros genéticos, valores genéticos e ganhos genéticos pelo procedimento da amostragem Gibbs podem ser mais precisas do que pelo procedimento REML/BLUP, inclusive para amostras pequenas, comuns em experimentos de campo. Assim, acredita-se haver urgente necessidade de assimilação e avaliação desta estratégia de análise nos experimentos de plantas. Pretende-se, dessa forma, desenvolver o processo de seleção entre e dentro de famílias de cana de açúcar usando a análise bayesiana.

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

Desenvolver um algoritmo computacional para seleção entre e dentro de famílias de cana de açúcar sob o enfoque bayesiano.

### **1.2.2 Objetivos Específicos**

- Comparar a análise de modelos mistos sob o enfoque bayesiano com a análise clássica via REML/BLUP
- Utilizar diferentes definições de distribuições *a priori* na modelagem;
- Avaliar a importância do uso de informação de parentesco nas análises.
- Disponibilizar o algoritmo para o usuário em geral sob a forma de conjunto de funções de fácil utilização

## 2 REFERENCIAL TEÓRICO

### 2.1 Modelos Mistos

Um modelo misto é aquele que apresenta tanto fatores de efeitos fixos como de efeitos aleatórios, além do erro experimental e da constante.

Um modelo misto pode ser expresso na forma geral (Henderson, 1984, citado por RESENDE, 2002)

$$y = X\beta + Zg + \varepsilon$$

em que:

y: vetor ( $n \times 1$ ) de observações;

X: matriz ( $n \times p$ ) de incidência dos efeitos fixos;

Z: matriz ( $n \times q$ ) de incidência dos efeitos aleatórios (genéticos ou ambientais);

$\beta$ : vetor ( $p \times 1$ ) de efeitos fixos a serem estimados;

g: vetor ( $g \times 1$ ) de efeitos aleatórios a serem preditos; e

$\varepsilon$ : vetor de erros aleatórios associados a cada observação.

Em geral, pressupõe-se que  $E(g) = 0$  e  $E(\varepsilon) = 0$ , e que:

$$V \begin{bmatrix} g \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix},$$

em que,  $\text{Var}(g) = G$ ,  $\text{Var}(\varepsilon) = R$ , e  $\text{Cov}(g, \varepsilon) = \text{Cov}(\varepsilon, g) = 0$ , sendo G e R matrizes simétricas positivas definidas.

Desta maneira,  $E(y) = X\beta$  e  $\text{Var}(y) = V = ZGZ' + R$ , bem como, condicionalmente aos efeitos aleatórios,  $E(y|g) = X\beta + Zg$  e  $\text{Var}(y|g) = R$ . Ainda,  $\text{Cov}(y, g) = ZG$ ; e  $\text{Cov}(y, \varepsilon) = R$ .

Obter a inversa de V exige grande demanda computacional. No entanto, uma alternativa para obtenção do BLUP já havia sido demonstrada por Henderson *et al.* (1959), citado por Resende (2002) e Littell *et al.* (2006), o que consiste na maximização da função densidade de probabilidade conjunta de y e g.

Assumindo que y, g e  $\varepsilon$  tenham distribuição normal multivariada, tem-se:

$$\begin{bmatrix} y \\ g \\ \varepsilon \end{bmatrix} \sim \text{NMV} \left\{ \begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} ZGZ' & ZG & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix} \right\}.$$

Observe que poderíamos construir uma partição da matriz  $G$ , de modo que uma de suas partes, digamos  $G_A$ , fosse igual a  $A\sigma_g^2$ , sendo  $A$  a matriz que indica o grau de associação genética entre os indivíduos (parentesco genético, ou matriz de relacionamento genético aditivo) e  $\sigma_g^2$  a variância genética associada. As outras partições da matriz  $G$ , em geral  $I\sigma_{g^*}^2$ , seriam referentes aos outros efeitos aleatórios. Dessa forma, informações de parentesco entre os indivíduos poderiam ser modeladas conjuntamente.

Para o caso geral, derivando a função de verossimilhança  $L$  em relação à  $\beta$  e  $g$  e tornando tais derivadas identicamente nulas, obtêm-se as seguintes equações de modelos mistos (EMM):

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}.$$

Da absorção de  $\hat{g}$  nas equações de  $\beta^0$  é possível a obtenção da solução para  $\beta^0$ , dada por:

$$\beta^0 = \left\{ X' \left[ R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \right] X \right\}^{-1} X' \left[ R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \right] y$$

$$\beta^0 = (X'V^{-1}X)^{-1}X'V^{-1}y$$

$$\text{Em que, } R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} = V^{-1}.$$

Tomando a segunda equação das EMM, tem-se o melhor preditor linear não-viesado (BLUP) de  $g$ , dado por:

$$\hat{g} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\beta^0).$$

Se  $G$  e  $R$  são conhecidas,  $X\beta^0$  é o melhor estimador linear não-viesado (BLUE) de  $X\beta$ , e  $\hat{g}$  é o melhor preditor linear não-viesado (BLUP) de  $g$ .

O uso da metodologia de modelos mistos, para predição de valores genéticos, requer inicialmente o conhecimento dos componentes de variância associadas a G e R. Geralmente, estes são desconhecidos, e as predições não serão necessariamente as melhores, mas uma aproximação. Assim, uma determinação segura das estimativas dos componentes da variância é fundamental para o aprimoramento do processo de estimação, como o de predição, podendo resultar em alterações nas estimativas dos efeitos fixos (*Best Linear Unbiased Estimation* - BLUE), bem como nas predições dos efeitos aleatórios (*Best Linear Unbiased Prediction* - BLUP) (RESENDE, 2002).

Existem duas abordagens para estimar os componentes de variância: Clássica e Bayesiana.

### 2.1.1 Enfoque Clássico

Um motivo para fazer uso do modelo linear misto clássico é a possibilidade de se fazer predição de efeitos aleatórios, através do *BLUP* e obter estimativas *BLUE* para efeitos fixos. O procedimento ótimo de estimação de componentes de variância é o *REML* (máxima verossimilhança restrita).

O método de Máxima Verossimilhança Restrita ou Residual (*REML*) utiliza uma função de verossimilhança a partir dos resíduos completos estimados,  $y - X\beta^0$ , ou seja, é definida somente nos termos dos parâmetros de variância e covariância, e dessa forma suas estimativas tornam-se não viciadas (BROWNE, DRAPER, 2006).

Os estimadores iterativos dos componentes de variância,  $\sigma_g^2$  (variância genotípica) e  $\sigma_\varepsilon^2$  (variância residual) por *REML*, são normalmente obtidos pelo algoritmo *EM* (*Expectation Maximization*) que alterna os passos de esperança e maximização. Suas etapas são dadas pelas expressões a seguir:

$$\hat{\sigma}_\varepsilon^2 = \frac{y'y - \beta^0'X'y - \hat{g}'Z'y}{n - r(X)}$$

$$\hat{\sigma}_g^2 = \frac{\hat{\mathbf{g}}' \mathbf{A}^{-1} \hat{\mathbf{g}} - \text{Tr}(\mathbf{A}^{-1} \mathbf{C}_{22}) \hat{\sigma}_\varepsilon^2}{q}$$

em que  $r(X)$  é o posto da matriz  $X$ ;  $A$  é a matriz de parentesco genético;  $tr(.)$  o traço da matriz;  $C_{22}$  é obtido da parte inferior esquerda da inversa da matriz de Henderson dos coeficientes das equações do modelo misto, que é dada a seguir:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\hat{\sigma}_\varepsilon^2 \end{bmatrix}^{-1}$$

conforme Resende (2002).

Logo, a melhor estimação linear não-viesada (BLUE) é dada por:

$$\text{BLUE}(\beta) = (X'V^{-1}X)^{-1}X'V^{-1}y$$

E a melhor predição linear não-viesada (BLUP) é dada por:

$$E(g|Y) = \text{BLUP}(g) = GZ'V^{-1}(y - X\beta^0)$$

Sendo  $\hat{\beta}$  igual a  $\text{BLUE}(\beta)$  e  $GZ' = \text{Cov}(y, g)$ , segundo Henderson (1975).

Nota-se que para obter as soluções do BLUE e do BLUP por meio das EMM, necessita-se primeiramente de valores arbitrários iniciais para os componentes de variância para a realização do algoritmo EM.

### 2.1.2 Enfoque Bayesiano

A inferência bayesiana é uma alternativa ao método clássico descrito acima, para estimação dos componentes de variância e dos efeitos fixos, bem como para predição dos efeitos aleatórios do modelo (LINDLEY & SMITH, 1972; BOX & TIAO, 1973).

Dentro do contexto bayesiano, qualquer inferência em relação às quantidades desconhecidas do modelo (incluindo parâmetros e variáveis não observáveis) é obtida a partir da distribuição conjunta a *posteriori* de  $\beta$ ,  $g$ ,  $G$  e  $R$ , representada por  $p(\beta, g, G, R | y)$  (SORENSEN e GIANOLA, 2002), em que:

$$p(\beta, g, G, R|y) = \frac{p(y|\beta, g, R) p(\beta, G, g, R)}{p(y)}$$

Como toda inferência Bayesiana é condicional aos dados observados  $y$ ,  $p(y)$  na expressão acima é uma constante, de maneira que a distribuição *a posteriori* conjunta não normalizada pode ser expressa proporcionalmente a:

$$p(\beta, g, G, R | y) \propto p(y | \beta, g, R)p(\beta, g, G, R),$$

onde  $p(y | \beta, g, R)$  representa o modelo amostral (função de verossimilhança) que no nosso caso, como descrito acima, refere-se a uma distribuição normal com vetor de médias  $X\beta+Zg$  e matriz de variâncias e covariâncias  $R$ , e  $p(\beta, g, G, R)$  representa a distribuição *a priori* conjunta das quantidades desconhecidas do modelo.

A distribuição *a priori* deve ser determinada pelo pesquisador, utilizando todo o seu conhecimento sobre o parâmetro que se estuda antes de iniciar a amostragem. Este conhecimento também pode ser obtido a partir de distribuições *a posteriori* de estudos anteriores, dentre outras fontes ou da opinião de especialistas. Se não há conhecimento algum a respeito do parâmetro, pode-se utilizar uma distribuição *a priori* não-informativa, e assim as estimativas obtidas a partir da distribuição *a posteriori* serão apenas baseadas na função de verossimilhança. (LEOTTI, 2007)

Em geral, em modelos lineares mistos, assume-se que  $p(\beta, g, G, R) = p(\beta)p(g | G)p(G)p(R)$ , onde  $p(\beta)$  é uma distribuição não informativa (geralmente assumida como uniforme),  $p(g | G)$  é uma distribuição normal multivariada com vetor de médias zero e matriz de covariâncias  $G$ , e  $p(G)$  e  $p(R)$  assumem diferentes distribuições dependendo das suposições do modelo em relação aos componentes de variância. Por exemplo, uma distribuição Wishart invertida é assumida para uma matriz de covariâncias não estruturada, e uma distribuição qui-quadrado escalonada invertida é adotada quando um único

componente de variância define a estrutura de covariância, como por exemplo, se é assumido que  $R = I\sigma^2$ .

Integrando-se a distribuição *a posteriori* conjunta  $p(\beta, g, G, R | y)$  obtêm-se as distribuições marginais de interesse (os limites de integração são referentes aos parâmetros que não são de interesse), a partir das quais podem ser obtidas medidas de posição (tais como médias, modas ou medianas *a posteriori*), bem como medidas de variabilidade e intervalos de probabilidade para os parâmetros de interesse. No caso dos valores genéticos, seus efeitos serão avaliados pelo valor da média *a posteriori*.

A construção de um intervalo de credibilidade em que a densidade de probabilidade dos valores se enquadram em sua maioria entre os limites definidos e a probabilidade do verdadeiro valor do parâmetro estar entre estes limites ser igual a  $1 - \alpha$  é denominado HPD (“Highest Probability Density”) (SMITH, 2007).

Existe também o intervalo de credibilidade central que pode ser obtido diretamente da distribuição *a posteriori* a partir dos quantis de interesse  $\left(\frac{\alpha}{2} \text{ e } 1 - \frac{\alpha}{2}\right)$  (LEOTTI, 2007).

Qualquer conjunto de dados pode ser analisado utilizando-se enfoque bayesiano, entretanto, modelos com muitos parâmetros necessitam de cálculos impossíveis de serem realizados, o que tornou o uso da inferência bayesiana um obstáculo durante muitos anos.

Com o avanço dos métodos computacionais o uso da inferência bayesiana tornou-se factível em muitos problemas estatísticos. Um dos grandes ganhos conforme Sorensen e Gianola (2002) foi a possível aproximação das integrais por métodos numéricos ou de Monte Carlo.

## **2.2 Método de Monte Carlo**

Os métodos de Monte Carlo referem-se a processos de aproximação de valores esperados (integrais com respeito a uma distribuição de probabilidade),

por meio de uma amostra, podendo ser referidos também como um caso especial de simulação de um processo estocástico (RESENDE, 2002).

Dentro dos métodos de Monte Carlo, a amostragem de Gibbs se apresenta como uma ferramenta bastante útil e eficiente que é indicada quando as distribuições condicionais completas são identificáveis (NOGUEIRA, 2003)

A distribuição de Gibbs é uma importante, e relativamente “velha” ferramenta da física usada principalmente em restauração de imagens. A origem do Amostrador de Gibbs foi através de Geman e Geman (1984) quando pela primeira vez utilizaram a simulação Monte Carlo via cadeias de Markov (geração de valores) dependentes do “estado” anterior para gerar amostras da distribuição de Gibbs, por isso essa denominação. Entretanto somente com o trabalho de Gelfand e Smith (1990) que pela primeira vez a técnica foi aplicada para gerar valores de distribuição Marginal a posteriori.

O amostrador de Gibbs se baseia na geração de uma cadeia de Markov a partir da amostragem alternada de cada distribuição condicional obtida da distribuição  $p(\beta, g, G, R | y)$ . Após um período de aquecimento da cadeia (*burn-in period*) para convergência da mesma, cada ciclo da cadeia fornece uma amostra aleatória da distribuição a *posteriori* conjunta. A partir destas amostras pode-se então aproximarem-se as medidas a *posteriori* de interesse, bem como estudar a distribuição marginal de cada parâmetro do modelo (GILKS *et al.*, 1995).

Os passos para a aplicação do Amostrador de Gibbs são descritos a seguir:

1°. Inicialize a cadeia e assuma valores iniciais, em que,  $\sigma_g^2$  é a variância genética aditiva e  $\sigma_\varepsilon^2$  a variância residual

$$\theta^0 = \{\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_n^{(0)}, g_1^{(0)}, g_2^{(0)}, \dots, g_n^{(0)}, \sigma_g^{(0)}, \sigma_\varepsilon^{(0)}\}$$

2°. Obtenha os próximos estados da cadeia,  $\theta^{(t+1)} = \{\beta_1^{(t+1)}, \beta_2^{(t+1)}, \dots, \beta_n^{(t+1)}, g_1^{(t+1)}, g_2^{(t+1)}, \dots, g_n^{(t+1)}, \sigma_g^{(t+1)}, \sigma_\varepsilon^{(t+1)}\}$  por meio das sucessivas “gerações de valores” das distribuições condicionais completas a posteriori.

$$\beta_1^{(t+1)} \sim P(\beta_1 / \beta_2^{(t)}, \beta_3^{(t)}, \dots, \beta_n^{(t)}, g_1^{(t)}, g_2^{(t)}, \dots, g_n^{(t)}, \sigma_g^{(t)}, \sigma_\varepsilon^{(t)})$$

$$\beta_2^{(t+1)} \sim P\left(\beta_2 / \beta_1^{(t+1)}, \beta_3^{(t)}, \dots, \beta_n^{(t)}, g_1^{(t)}, g_2^{(t)}, \dots, g_n^{(t)}, \sigma_g^{(t)}, \sigma_\varepsilon^{(t)}\right)$$

$$\vdots$$

$$\sigma_g^{(t+1)} \sim P\left(\sigma_a / \beta_1^{(t+1)}, \beta_3^{(t+1)}, \dots, \beta_n^{(t+1)}, g_1^{(t+1)}, g_2^{(t+1)}, \dots, g_n^{(t+1)}, \sigma_\varepsilon^{(t)}\right)$$

$$\sigma_\varepsilon^{(t+1)} \sim P\left(\sigma_e / \beta_1^{(t+1)}, \beta_3^{(t+1)}, \dots, \beta_n^{(t+1)}, g_1^{(t+1)}, g_2^{(t+1)}, \dots, g_n^{(t+1)}, \sigma_g^{(t+1)}\right)$$

3º. Incrementar t

4º. Voltar ao passo 2 e continuar as iterações até um tamanho  $m$  que garanta a convergência do algoritmo

Ao final obtêm-se amostras das distribuições marginais a posteriori de cada parâmetro.

Dentre os métodos MCMC (*Markov Chain Monte Carlo*), o amostrador de Gibbs é bastante usado e pode ser executado através da realização de algoritmo no programa R.

### 2.3 Diagnóstico de Convergência

Como o Amostrador de Gibbs define uma cadeia de Markov, a amostra gerada através dele não é independente. Cada novo valor depende dos valores gerados na iteração anterior e dessa forma, a sequência apresenta uma autocorrelação, o que torna a estimativa do erro padrão da média viesado. Um método afim de tentar eliminar a autocorrelação e obter uma amostra de tamanho ideal da distribuição de interesse foi proposto por Raftery e Lewis (1992) e a monitoração da convergência das sequências nas proximidades do tamanho indicado para a amostra é bem verificado por Geweke (1992). (NOGUEIRA, SÁFADI, FERREIRA, 2004).

Existem várias técnicas que auxiliam na identificação e monitoramento da convergência da sequência. Maiores detalhes podem ser vistos em Nogueira (2004). Neste trabalho falaremos somente das duas citadas acima.

### 2.3.1 Critério de Raftery e Lewis (1992)

É o método que estima quantas iterações são necessárias para que o Amostrador de Gibbs apresente convergência à distribuição estacionária, propondo especificações de quantidades necessárias para isto. As estimativas fornecidas são o número de iterações que devem ser descartadas  $B$  (Burn-in), a distância mínima entre uma iteração e outra para se obter uma subamostra aproximadamente independente  $K$  (thin), o número de iterações que devem ser computadas  $N$ , o número de amostras independentes necessárias para estimar a probabilidade a posteriori dentro do grau de precisão desejado e a probabilidade de cobertura ("Lower Bound"), por fim é fornecido o número total de iterações dividido pelo limite inferior que mede o aumento do tamanho da amostra devido a autocorrelação (Fator dependência). Se o fator de dependência for maior que 5, pode-se concluir que a cadeia não atingiu convergência, necessitando reparametrizar o modelo em estudo. Portanto, deve-se tomar os maiores valores de  $B$ ,  $N$  e  $K$ , pois a convergência só se dará quando as cadeias de todas as variáveis convergirem.

Os valores serão calculados mediante especificações para garantir que o quantil  $u$  de uma determinada função  $f(\theta)$  seja estimado com precisão pré-definida.

Considere a estimativa da probabilidade  $P(\theta \leq u | y)$  da posterior de um parâmetro de modelo com um erro  $\epsilon$  e probabilidade  $\gamma$ . Os passos do método de Raftery e Lewis são os seguintes:

- i) Calcula-se  $\theta^i$  e o valor de  $Z_i = I(f(\theta^i) \leq u)$  para cada iteração, em que  $I(\cdot)$  é a função indicadora e  $Z_i$  é uma cadeia binária de (0 ou 1) e  $f(\theta^i)$  é o valor gerado da  $i$ -ésima iteração.
- ii) A partir da seleção de observações  $Z_i$  considerando saltos de  $K$  (menor valor para o qual a dependência em apenas um estágio é preferida) unidades, segue um novo processo  $Z_i^K = Z_i + (i - 1)K$ .
- iii) Admite que a matriz de transição de  $Z_i^K = \begin{bmatrix} 1 - \alpha & \beta \\ \beta & 1 - \alpha \end{bmatrix}$  em que a distribuição de equilíbrio é dada por  $\pi = (\pi_0, \pi_1)$ ,  $\pi_0 = P(\theta \leq$

$u|y) = \frac{\beta}{\alpha+\beta}$  e  $\pi_1 = 1 - \pi_0 = \frac{\alpha}{\alpha+\beta}$ . Pode-se mostrar que o período de aquecimento é  $B = mK$ , sendo,  $m = \frac{\log\left(\frac{\epsilon(\alpha+\beta)}{\max(\alpha,\beta)}\right)}{\log(1-\alpha-\beta)}$ , em que  $\epsilon$  é a distância requerida entre  $P(Z_B^K = l|Z_0^K = j)$  e a probabilidade de equilíbrio ( $l, j = 0, 1$ ).

- iv) Usando o fato de que a média amostral de  $\theta$  tem distribuição assintótica normal, pode-se obter o tamanho da amostra necessário para que se estime  $P(\theta \leq u|y)$  com um erro máximo igual a  $\epsilon$  com probabilidade  $\gamma$ . A distribuição assintótica de  $\bar{Z}^K = \frac{1}{N} \sum_{i=1}^N Z_i^K$  é normal com média  $p$  e variância  $\frac{\alpha\beta(2-\alpha-\beta)}{N(\alpha+\beta)^3}$  e daí tem-se que

$$N = \frac{\alpha\beta(2-\alpha-\beta)}{N(\alpha+\beta)^3} \left[ \frac{\Phi\left(\frac{1}{2}(1+i)\right)}{r} \right]^2 \text{ em que } \Phi \text{ é a função distribuição}$$

acumulada da distribuição normal padrão.

Para implementar o método é necessário simular o processo e estimar os valores de  $\alpha$  e  $\beta$ .

### 2.3.2 Critério de Geweke (1992)

É um diagnóstico para a ausência de convergência do Amostrador de Gibbs, baseado na técnica padrão de análise espectral que promove um erro padrão para a aproximação  $E(g(\theta))$  das média amostrais correspondentes, o que leva em conta qualquer autocorrelação em que  $g(\theta)$  é uma função real.

Considerando a função real  $g(\theta)$  e sua trajetória  $g^{(1)}, g^{(2)}, \dots$  construída a partir de  $g^{(i)} = g(\theta^{(i)})$ ,  $i = 1, 2, \dots, n$ , essa trajetória define uma série temporal.

A média desta série é dada por:  $\bar{g}_i = \frac{\sum_{i=1}^n g_i}{n}$

Ela fornece um estimador não-viesado de  $E[g(\theta)]$  cuja variância assintótica é dada por  $\frac{S_g(0)}{n}$ , em que  $S_g(w)$  é a densidade espectral da série  $g$  de frequência.

Se a sequência  $g^{(i)}$  for estacionária, a diferença padronizada entre as médias é dada por:

$$\frac{\bar{g}_a - \bar{g}_b}{\sqrt{\left(\frac{\hat{S}_g^a(0)}{n_a}\right)\left(\frac{\hat{S}_g^b(0)}{n_b}\right)}} \sim N(0,1)$$

em que  $\bar{g}_a$  é a média das primeiras iterações e  $\bar{g}_b$  é a média das últimas iterações.

A regra de decisão do critério é baseada na estimativa do p-valor. Se o p-valor for menor que um nível de significância adotado pelo pesquisador, conclui-se que a cadeia não atingiu convergência.

## 2.4 BLUPIS

A estratégia de seleção ótima em cana de açúcar, segundo Resende (2002), seria através da predição de valores genotípicos usando o BLUP (Best Linear Unbiased Predictor) individual (BLUPI). Este procedimento usaria, simultaneamente, a informação de família e de indivíduos para a seleção. No entanto esse método dificilmente é usado em programas de melhoramento devido a problemas operacionais relacionados à obtenção dos dados em nível de planta.

Para contornar esta situação Resende & Barbosa (2006) propuseram a seleção via BLUP (*Best Linear Unbiased Predictor*) individual simulado (BLUPIS), que se baseia nos efeitos genotípicos das famílias avaliadas a campo, isto é, quando progênies inteiras são selecionadas ou rejeitadas como unidade, de acordo com seu valor genotípico médio.

A eficiência da seleção de famílias é baseada no fato de que os desvios dos efeitos ambientais dos indivíduos tendem a se anular e é preferida quando o caráter selecionado apresenta baixa herdabilidade (FALCONER & MACKAY, 1996).

O método BLUPIS, descrito por Resende e Barbosa (2006), é considerado a melhor estratégia de seleção na etapa inicial de progênies, pois consiste em determinar de forma dinâmica o número de famílias a contribuir com indivíduos

selecionados e o número de indivíduos a serem selecionados em cada família, sem que haja avaliação individual destes. O uso deste método é ideal para ser empregado em cana-de-açúcar, onde as parcelas experimentais são colhidas em sua totalidade.

A expressão que determina o número  $n_k$  de indivíduos a serem selecionados em cada família  $k$  é dada por:  $n_k = (\hat{g}_k / \hat{g}_j) n_j$ , em que:  $\hat{g}_j$  refere-se ao efeito genotípico da melhor família;  $\hat{g}_k$  refere-se ao efeito genotípico da  $k$ -ésima família; e,  $n_j$  equivale ao número de indivíduos selecionados na melhor família. Segundo Resende e Barbosa (2005) a escolha de  $n_j$  pode ser igual a 50. Os autores acreditam que 50 indivíduos selecionados dentro da melhor família são suficientes para conter o melhor indivíduo da progênie para o caráter produtividade, pois este número conseguiria níveis próximos de 98% da representatividade máxima da família.

Este método, em sua forma original, segundo os autores, elimina automaticamente as famílias com efeito genotípico negativo, ou seja, aquelas abaixo da média geral do experimento, o que ocasiona um menor número de melhores indivíduos, permitindo assim, aumentar a eficiência do processo seletivo e diminuindo custo no programa de melhoramento.

## 2.5 Informação de Parentesco

É a informação obtida por meio de elementos que descrevem os coeficientes das covariâncias entre os efeitos aleatórios genéticos, estes conhecido também como coeficiente de parentesco de Malécot (KEMPTHORNE, 1973).

A matriz de parentesco é denotada por  $A$ , e segue o esquema abaixo, para o caso de três indivíduos:

$$A = \begin{bmatrix} 2f_{11} & 2f_{12} & 2f_{13} \\ 2f_{21} & 2f_{22} & 2f_{23} \\ 2f_{31} & 2f_{32} & 2f_{33} \end{bmatrix}$$

em que  $f_{(x,y)}$  é o coeficiente de parentesco de Malécot entre os indivíduos  $x$  e  $y$ .

O coeficiente de parentesco de Malécot entre  $x$  e  $y$  corresponde à probabilidade, de dois alelos, tomado ao acaso em  $x$  ser idêntico por descendência a um alelo tomado ao acaso em  $y$  (Souza Júnior, 1989).

Como as características das plantas sofrem muita influência do meio ambiente, o histórico familiar diz muito sobre sua genética.

A alta acurácia propiciada pelos modelos mistos se deve ao fato de levarem em conta associações genéticas existentes entre os indivíduos, em razão do parentesco entre elas, quantificadas por informações de genealogia.

O uso da matriz de parentesco melhora a precisão das estimativas dos valores genéticos e quase sempre aumenta a correlação entre os valores genéticos dos indivíduos aparentados (Henderson, 1976).

A obtenção da análise dos coeficientes de parentesco através do pacote estatístico R pode ser mais bem entendida por meio do trabalho de PETERNELLI et. al. (2009).

### 3 REFERÊNCIAS

BOX, G.E.P.; TIAO, G.C. **Bayesian inference in statistical analysis**. Reading: Addison-Wesley Publ. Co. 1973. p.588.

BROWNE, W.J., DRAPER D. Implementation and performances issue in the Bayesian and likelihood fitting of multilevel models. **Bayesian analysis**, 2006, 1(3): 473-514.

CESNIK, R. **Melhoramento da cana-de-açúcar**. Brasília: Embrapa Informações tecnológicas, 2004, p.307.

COGO, Carlos. **Cana-de-açúcar**. Disponível em:  
[http://www.deere.com.br/pt\\_BR/ag/veja\\_mais/info\\_mercado/sugar\\_cane.html](http://www.deere.com.br/pt_BR/ag/veja_mais/info_mercado/sugar_cane.html).  
Acesso em: 15 de jun. 2011.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4. Ed. London: Longman, 1996. p.464.

GELFAND A.E, SMITH A.F.M. Sampling-Based Approaches to Calculating Marginal Densities. **Journal of the American Statistical Association**, 1990, Jun; 85(410): 398-409.

GEMAN S., GEMAN D., Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. **Ieee Transactions on Pattern Analysis and Machine intelligence**, 1984; 6(6):721-41.

GEWEKE J. chapter Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. **Bayesian Statistics**. Oxford University Press, New York. v.4, 1992.

HENDERSON, C. R. et al. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v.15, 1959 p.192.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v. 31, n. 3, 1975, p. 423-447.

HENDERSON, C. R., A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. **Biometrics**, 1976, 32: 69-83.

HENDERSON, C. R. **Applications of linear models in animal breeding**. Ontario: University of Guelph, 1984, p. 462.

KEMPTHORNE, O. **An introduction to genetic statistics**. Ames: Iowa State University Press, 1973, p 545.

KIMBENG, C.A.; COX, M.C. Early generation selection of sugarcane families and clones in australia: a review. **Journal American Society of sugarcane Technologists**, v.23, 2003, p.20-39.

LEOTTI, V.B. **Comparação Via Simulação dos Estimadores Clássicos e Bayesianos no Modelo de Coeficientes Aleatórios para Dados Longitudinais**. 2007. 99 f. Dissertação (Programa de Pós-Graduação em Epidemiologia) - Universidade Federal do Rio Grande do Sul, RS, 2007.

LINDLEY, D.V.; SMITH, A.F.M. Bayes estimates for the linear model. **Journal of the Royal Statistical Society: Series B**. v.34, 1972, p.1-41.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. **Melhoramento da cana-de-açúcar**. In: Borém, A. (2ª ed.). Melhoramento de espécies cultivadas. UFV: Viçosa, 2005. p.225-274.

NOGUEIRA, D. A. et al. Análise clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Cienc. Agrotec.**, Lavras Edição Especial, dez., 2003, p. 1614-1624.

OLIVEIRA, R.A. et al. Seleção de famílias de cana-de-açúcar via modelos mistos. **Scientia Agrária**, Curitiba, v.9, n.3, 2008, p.269-274.

PETERNELLI, L.A.; FERREIRA, F.M.; ROCHA, R.B.; BARROS, W.S.; BARBOSA, M.H.P. Análise dos coeficientes de endogamia e de parentesco para qualquer nível de ploidia usando o pacote estatístico R. **Bragantia**, v.4. 2009.

RAFTERY, A.L.; LEWIS, S. chapter How Many Iterations in the Gibbs Sampler? **Bayesian Statistics**. Oxford University Press, New York. v.4, 1992.

R DEVELOPMENT CORE TEAM. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Áustria, 2011. Disponível:<<http://www.R-project.org>>.

RESENDE, M.D.V. de et al. Mixed linear models analysis using Bayesian inference. **Rev. Mat. Estat.**, São Paulo, 2001, 19: 41-70.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica. 2002, p.975

RESENDE, M. D. V. de; BARBOSA M.E.P. Selection via simulated individual BLUP based on family genotypic effects in sugarcane. **Pesq. agropec. bras.**, Brasília, v.41, n.3, mar. 2006, p.421-429.

SMITH. B. J. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. **Journal of Statistical Software**: Issue 11. v.21, 2007, p.1-41.

SORENSEN, D; GIANOLA, D. **Likelihood, Bayesian and MCMC Methods in Quantitative Genetics**. Springer. 2002. p.760.

SOUZA JUNIOR, C.L. **Componentes de variância genética e suas implicações no melhoramento vegetal**. Piracicaba: FEALQ, 1989. p.134 .

## CAPÍTULO I - ANÁLISE DE DADOS REAIS

### 1 INTRODUÇÃO

A composição genética de diversas culturas de plantas é consumada a partir de um processo de melhoramento a que elas são submetidas durante décadas. Todo fenomenal progresso genético de um vasto número de espécies é concluído com a ajuda de melhoristas e o desenvolvimento de técnicas estatísticas apropriadas para otimizar o processo de seleção. A cana-de-açúcar (*Saccharum spp.*) é uma delas, que devido a sua grande produção em açúcar e álcool, é bastante estudada em programas de melhoramento.

A metodologia bastante usada em diversas áreas do conhecimento, sendo de grande aplicabilidade na agricultura para análise de dados para seleção, é o modelo linear misto (LACHOS, 2005). Esse modelo compreende tanto efeitos fixos como aleatórios em sua formulação e assume distribuição multivariada tanto para os efeitos aleatórios como para os erros.

A estratégia ótima de seleção de indivíduos nas fases iniciais do melhoramento de cana-de-açúcar, que contempla simultaneamente as informações de família e de indivíduos para a seleção, seria através dos valores genotípicos preditos pelo BLUP (*Best Linear Unbiased Prediction*) individual (BLUPI) (RESENDE, 2002a). No entanto, há dificuldades práticas em se obter dados de plantas individuais.

Segundo Resende e Barbosa (2006) a variação genética dentro de famílias pode inferir em potencial de certos cruzamentos para gerar indivíduos superiores, mas sua estimativa apresenta erros maiores que a variação genética entre famílias, além de aumentar os custos.

No entanto, a seleção individual de clones (seleção massal) aplicada nos estágios iniciais dos programas, apesar de ser usualmente praticado (MATSUOKA *et al.*, 2005) tem se mostrado ineficiente (KIMBENG & COX, 2003). Esses autores discutem que a seleção de famílias seguida da seleção

individual de clones, apresenta maior ganho genético do que a simples seleção de clones ou de famílias. Além disso, a economia de recursos seria muito maior.

Um método que consiste determinar de forma dinâmica o número de indivíduos a serem selecionados em cada família, sem que haja avaliação individual destes foi descrito por Resende e Barbosa (2006) e denominado BLUP Individual Simulado (BLUPIS).

Este método é bastante usado na literatura via enfoque clássico (REML/BLUP), pois como diz Gianola (1996) o enfoque bayesiano é mais trabalhoso e sua obtenção pela forma analítica é bastante difícil, o que dificulta seu uso.

Contudo, a fim de obter os efeitos genotípicos de cana-de-açúcar por ambos os enfoques e poder compará-los, utilizou-se o programa Selegen (Resende, 2002b)) para obter a análise clássica e foi construído um algoritmo no software livre R para a obtenção da análise bayesiana.

## **2 MATERIAL E MÉTODOS**

### **2.1 Detalhes experimentais**

As famílias utilizadas nesta pesquisa foram provenientes de cruzamentos realizados na Estação Experimental da Serra do Ouro, pertencente à Universidade Federal de Alagoas, localizada no Município de Murici, Estado de Alagoas, no ano de 2006. Os cruzamentos realizados foram controlados, denominados biparentais ou cruzamentos simples. Para o presente estudo, foram utilizadas 22 famílias.

As sementes originadas destes cruzamentos foram coletadas e acondicionadas em embalagens apropriadas, sendo, posteriormente, enviadas ao Centro de Pesquisa e Melhoramento de Cana-de-açúcar (CECA), da Universidade Federal de Viçosa e em 2007 foi feito o experimento em delineamento em blocos casualizados (DBC) com cinco repetições. Cada família foi constituída por 100 plantas, ao todo distribuída nos 5 blocos. Os dados de cana planta foram coletados em junho 2008. A variável estudada neste trabalho é referente à tonelada de cana

por hectare (TCH). Para avaliar a eficiência da análise bayesiana frente a análise REML/BLUP (máxima verossimilhança restrita / melhor predição linear não viciada), estas foram feitas em DBC com 5 repetições.

A variável TCH real foi calculada utilizando a fórmula:

$$TCH = \frac{MTP \times 10}{tp}$$

em que, MTP é a massa total da família em kg, e tp é o tamanho da família em m<sup>2</sup>.

Utilizando a variável TCH, foi possível avaliar o desempenho do algoritmo construído.

## 2.2 Procedimentos utilizados

### 2.2.1 Modelos Mistos

As equações de modelos mistos foram utilizadas para facilitar o cálculo dos efeitos genotípicos das famílias.

O modelo estatístico utilizado é expresso pela equação  $y = X\beta + Zg + \varepsilon$ , em que y é o vetor (n×1) de observações, X é a matriz (n×p) de incidência dos efeitos fixos, Z é a matriz (n×q) de incidência dos efeitos aleatórios (genéticos ou ambientais), β é o vetor (p×1) de efeitos fixos a serem estimados, g é o vetor (q×1) de efeitos aleatórios a serem preditos e ε é o vetor de erros aleatórios associados a cada observação.

A obtenção da estimativa de β, demonstrada por Henderson *et. al.* (1959), é dada por:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

E a predição de g demonstrada por Henderson (1973) é dada por:

$$\hat{g} = GZ'V^{-1}(y - X\hat{\beta})$$

Em que  $V = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'$  é a matriz de covariância de y,  $G = A\sigma_g^2$ .

Admitiu-se neste presente trabalho que os erros são não correlacionados, isto é,

$\text{Cov}(y, \varepsilon) = R = I\sigma_\varepsilon^2$ , em que  $I$  é a matriz identidade.

Para o modelo adotado foi considerado:

$$g \sim N(0, A\sigma_a^2) \quad \text{e} \quad \varepsilon \sim N(0, I\sigma_\varepsilon^2)$$

em que  $\sigma_g^2$  é a variância genética aditiva e  $\sigma_\varepsilon^2$  a variância residual.

### 2.2.2 Método REML/BLUP

A análise de modelos mistos via inferência clássica, é realizada quando se conhece os componentes de  $G$ , ou então, quando estes são estimados e substituídos nas equações do modelo misto, obtendo aproximações para os melhores estimadores e preditores lineares de  $\beta$  e de  $g$ .

O método de estimação dos componentes de variância adotado foi o de Máxima Verossimilhança Restrita, ou Residual (REML). Este método utiliza uma função de verossimilhança que não envolve  $\beta$ , ou seja, é definida somente em termos dos parâmetros de variância e covariância (LEOTTI, 2007). Toda a análise foi feita por meio do software Selegen-REML/BLUP (Sistema Estatístico e Seleção Genética Computadorizada via Modelos Lineares Mistos) (RESENDE, 2002b).

Os intervalos de confiança para os valores genotípicos das famílias pelo método REML/BLUP foram calculados em conjunto com a estimativa SEP (desvio padrão do valor genotípico predito) por meio da expressão (RESENDE, 2002b):

$$I.C. = g \pm t SEP$$

Considerando  $t=1,96$  que é o valor tabelado da distribuição  $t$  de Student de acordo com o nível de significância de 5% .

Os intervalos de confiança referentes às estimativas dos componentes da variância foram calculados de acordo com a metodologia apresentada por Barbin (1993), sendo o intervalo da variância genotípica realizado por meio do seguinte cálculo:

$$\frac{gl_g \hat{\sigma}_g^2}{x_{gl_g, 0.975}^2} \leq \sigma_g^2 \leq \frac{gl_g \hat{\sigma}_g^2}{x_{gl_g, 0.025}^2}$$

em que,  $gl_g = 7,32$  é o grau de liberdade da variância genotípica, e  $x^2$  é a distribuição qui-quadrado.

E o intervalo da variância residual é dado através do cálculo aproximado pela distribuição t:

$$\hat{\sigma}_\varepsilon^2 \pm t_{(gl, 0.025)} s(\hat{\sigma}_\varepsilon^2)$$

$$\text{em que, } s(\hat{\sigma}_\varepsilon^2) = \sqrt{\frac{2(\hat{\sigma}_\varepsilon^2)^2}{gl+2}} \text{ e } t \text{ é o valor tabelado com 5\% de}$$

significância e  $gl$  é o grau de liberdade do erro= 21.

As estimativas dos intervalos de confiança da herdabilidade entre famílias foi realizada através da expressão apresentada por Knapp et al. (1985) que é definida como  $1 - \left(\frac{\theta_1}{\theta_2}\right)$ , e os intervalos estimados pela equação:

$$P \left\{ 1 - \left[ \left( \frac{M_1}{M_2} \right) F_{(1-\frac{\alpha}{2}; gl_2, gl_1)} \right]^{-1} \leq 1 - \left( \frac{\theta_1}{\theta_2} \right) \leq 1 - \left[ \left( \frac{M_1}{M_2} \right) F_{(\frac{\alpha}{2}; gl_2, gl_1)} \right] \right\} = 1 - \alpha$$

em que,  $M_1$  foi o valor estimado para o quadrado médio do tratamento e  $M_2$  foi o valor estimado para o quadrado médio do erro.  $\theta_1$  e  $\theta_2$  foram os reais valores dos quadrados médios do tratamento e do erro, respectivamente,  $\alpha$  é o nível de significância e  $F$  é o valor tabelado da distribuição F com  $\alpha= 5\%$  e grau de liberdade do  $M2=M1=21$ .

Para ter conhecimento sobre os graus de liberdade da variância genotípica, foi necessário saber a esperança de quadrado médio da família por meio de um experimento instalado segundo DBC. Considerou-se o seguinte modelo estatístico para a realização desta análise.

$$Y_{ij} = \mu + G_i + B_j + \varepsilon_{ij}$$

em que:

$Y_{ij}$ : observação obtida do i-ésimo tratamento avaliada no j-ésimo bloco;

$\mu$ : média geral do experimento;

$G_i$ : efeito aleatório da  $i$ -ésima família;

$B_j$ : efeito fixo do  $j$ -ésimo bloco;

$\varepsilon_{ij}$ : erro aleatório.

O esquema das esperanças dos quadrados médios, para o modelo estatístico acima está representado na tabela abaixo.

Tabela 1 - Esquema da análise de variância para o delineamento em blocos casualizados completos mostrando fonte de variação (FV), graus de liberdade (GL) e esperanças de quadrado médio (E(QM)).

FV	GL	QM	E(QM)
Blocos	$b - 1$	QMB	$\sigma_\varepsilon^2 + g\sigma_b^2$
Famílias	$g - 1$	QMF	$\sigma_\varepsilon^2 + b\sigma_g^2$
Resíduo	$(b - 1)(g - 1)$	QMR	$\sigma_\varepsilon^2$

Os estimadores utilizados para estes parâmetros são descritos a seguir:

$\sigma_\varepsilon^2$  (variância ambiental entre médias das famílias) = QMR

$\sigma_g^2$  (variância genotípica das famílias) =  $\frac{QMF - QMR}{b}$

$\sigma_b^2$  (variância genotípica dos blocos) =  $\frac{QMB - QMR}{g}$

Um modo de se obter uma aproximação dos g.l da variância genotípica ( $n_g$ ) é por meio da fórmula de Satterthwaite (BARBIN, 1993).

$$n_g = \frac{(QMF - QMR)^2}{\frac{(QMF)^2}{g - 1} + \frac{(QMR)^2}{(b - 1)(g - 1)}}$$

### 2.2.3 Método Bayesiano

O método BLUPIS (BLUP Individual Simulado) foi descrito por Resende e Barbosa (2006) e consiste em determinar de forma dinâmica o número de indivíduos a serem selecionados em cada família, sem que haja avaliação individual destes. A análise foi baseada no modelo misto em que para sua

realização com enfoque bayesiano foi necessário alguns elementos definidos a seguir.

Considera-se, inicialmente, segundo Sorensen (1996), a distribuição condicional dos dados ( $y$ ) como uma normal multivariada:

$$y | \beta, g, \sigma_\varepsilon^2 \sim N(X\beta + Zg, I \sigma_\varepsilon^2) \text{ em que } I \text{ é a matriz de identidade.}$$

Logo, a função de verossimilhança é dada por:

$$L(y | \beta, g, \sigma_\varepsilon^2) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(y - X\beta - Zg)'(y - X\beta - Zg)\right\}.$$

O segundo passo foi definido a distribuição a priori dos parâmetros do modelo.

Para o efeito fixo, foi definida uma priori do tipo constante, já que a noção do conhecimento a priori é vago para  $\beta$ .

A distribuição *a priori* dos componentes de variância  $\sigma_g^2$  (variância genética aditiva) e  $\sigma_\varepsilon^2$  (variância residual) foi especificada como uma distribuição qui-quadrado escalonada invertida, da forma:

$$p(\sigma_i^2 | v_i, S_i^2) \propto (\sigma_i^2)^{-(\frac{v_i}{2}+1)} \exp\left[\frac{-v_i S_i^2}{2\sigma_i^2}\right] \quad i = \varepsilon, g$$

Quando foi assumida uma distribuição *a priori* informativa, substituiu-se em  $v$  os graus de liberdade da distribuição qui-quadrado e em  $S_i^2$ , o valor inicial da variância.

Para assumir uma distribuição *a priori* não-informativa, a qui-quadrado foi reduzida em uma uniforme imprópria, substituindo-se  $v = -2$  e  $S_i^2 = 0$ .

Após definidas as distribuições acima, pôde-se escrever a distribuição conjunta *a posteriori* dos parâmetros do modelo,  $p(\beta, g, \sigma_g^2, \sigma_\varepsilon^2 / y)$ .

$$\begin{aligned} p(\beta, g, \sigma_g^2, \sigma_\varepsilon^2 | y) &\propto p(\beta, g, \sigma_g^2, \sigma_\varepsilon^2) p(y | \beta, g, \sigma_g^2, \sigma_\varepsilon^2). \\ p(\beta, g, \sigma_g^2, \sigma_\varepsilon^2 / y) &\propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{(n+v_\varepsilon+1)}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}[(y - X\beta - Zg)'(y - X\beta - Zg) \right. \\ &\quad \left. + v_\varepsilon S_\varepsilon^2]\right\} \left(\frac{1}{\sigma_g^2}\right)^{\frac{(q+v_g+1)}{2}} \exp\left\{-\frac{1}{2\sigma_g^2}[g'A^{-1}g + v_g S_g^2]\right\}. \end{aligned}$$

Dentro do contexto bayesiano, qualquer inferência em relação às quantidades desconhecidas do modelo (incluindo parâmetros e variáveis não observáveis) é obtida a partir da distribuição conjunta *a posteriori* (SORENSEN e GIANOLA, 2002). Logo, a partir da distribuição conjunta obtiveram-se as distribuições *a posteriori* condicionais completas para cada parâmetro:

$$\beta_i | \beta_{-i}, g, \sigma_g^2, \sigma_\varepsilon^2, y \sim N(\hat{\beta}_i, (X_i' X_i)^{-1} \sigma_\varepsilon^2);$$

$$g_i | \beta, g_{-i}, \sigma_g^2, \sigma_\varepsilon^2, y \sim N(\hat{g}_i, (Z_i' Z_i + A_{i,i}^{-1} g)^{-1} \sigma_\varepsilon^2);$$

$$\sigma_g^2 | \beta, g, \sigma_\varepsilon^2, y \sim \chi^{-2}(q + v g), \text{ em que o parâmetro de escala é } g' A^{-1} g + v_g S_g^2;$$

$$\sigma_\varepsilon^2 | \beta, g, \sigma_g^2, y \sim \chi^{-2}(n + v \varepsilon), \text{ em que parâmetro de escala é } (y - X\beta - Zg)'(y - X\beta - Zg);$$

E as aproximações para as distribuições marginais foram feitas através da implementação do Amostrador de Gibbs.

$$\text{A herdabilidade, } h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\varepsilon^2}{r}}, \text{ foi obtida em cada iteração.}$$

Para a determinação do número  $n_k$  de indivíduos a serem selecionados em cada família k foi utilizado à expressão:  $n_k = (\hat{g}_k / \hat{g}_j) n_j$ , em que:  $\hat{g}_j$  refere-se ao efeito genotípico da melhor família;  $\hat{g}_k$  refere-se ao efeito genotípico da k-ésima família; e,  $n_j$  equivale ao número de indivíduos selecionados na melhor família (RESENDE & BARBOSA, 2006). Estes autores também comentam que este método elimina automaticamente as famílias com efeito genotípico negativo, ou seja, aquelas abaixo da média geral do experimento.

Os intervalos referentes às predições dos componentes da variância, herdabilidade e valores genotípicos das famílias pela análise bayesiana foram calculados através do intervalo de credibilidade central, que calcula os valores relativos aos quantis de 2,5 % e 97,5% da distribuição *a posteriori*  $\left(\frac{\alpha}{2} \text{ e } 1 - \frac{\alpha}{2}\right)$ , para os limites inferior e superior, respectivamente.

#### **2.2.4 Comparação entre os enfoques clássico e bayesiano na obtenção dos efeitos genéticos**

Os resultados obtidos pelo algoritmo no R utilizando o método bayesiano em cana planta foram analisados em relação a ausência ou presença de informação de parentesco e priori informativa como também comparados em relação ao método clássico (REML/BLUP) realizado com o auxílio do programa computacional Selegen-Reml-Blup (RESENDE, 2002b). As análises dos dados foram baseadas nas comparações entre os métodos utilizando 5 repetições. Foram verificados a variância genética, a variância residual, a herdabilidade e os efeitos genotípicos junto a seus intervalos, como também conferindo se as seleções entre e dentro de famílias foram coincidentes.

#### **2.2.5 Informação de Parentesco**

A informação de parentesco, baseada no coeficiente de parentesco de Malécot (KEMPTHORNE, 1973) foi utilizada a fim de verificar sua importância junto à análise. Para isto foi usado a informação do pedigree das famílias em estudo e estas foram realizadas de duas maneiras; sem informação de parentesco ou com informação de parentesco.

Dessa forma é verificado se a informação de parentesco, quando ausente ou presente, tem alguma influência na avaliação e identificação das famílias geneticamente superiores, em fase inicial de seleção, dentro do programa de melhoramento.

### **3 RESULTADOS E DISCUSSÕES**

O processo de análise de dados pelo método bayesiano foi realizado por meio da criação de um algoritmo junto ao programa R, que dessa forma pode construir de maneira satisfatória todos os passos para a aplicação do Amostrador de Gibbs.

As tabelas de 2 a 5 correspondem aos resultados das estimativas e predições de parâmetros de modelos mistos com enfoque Clássico (REML/BLUP) e com o enfoque bayesiano. Para a realização do BLUP foram executadas 70000 iterações, respeitando o *burn-in* (número de iterações iniciais descartadas) e *thin* (distância mínima entre uma iteração e outra para se obter uma subamostra independente) indicado pelo método de convergência de Raftery e Lewis e sua convergência foi alcançada sendo verificada pelo teste Geweke (1992). Ambos os testes foram obtidos por meio do pacote BOA (*Bayesian Output Analysis Program*) no programa R (SMITH, 2007).

Na Tabela 2 nota-se que os resultados das estimativas e predições de parâmetros de um modelo misto com enfoque Clássico usando o método REML/BLUP foi bem parecido em comparação ao enfoque bayesiano sem informação de parentesco e com a distribuição a priori não-informativa. Foi utilizado um burn-in de 25 e thin de 8 iterações e a convergência foi alcançada.

Na Tabela 3 nota-se que os resultados das estimativas e predições de parâmetros de um modelo misto com enfoque Clássico usando o método REML/BLUP foi bem parecido em comparação ao enfoque bayesiano sem informação de parentesco e com priori informativa. Utilizou-se um burn-in 8 e thin 4 iterações e a convergência foi alcançada.

Na Tabela 4 nota-se que estimativas e predições de parâmetros de um modelo misto com enfoque Clássico usando o método REML/BLUP tiveram resultados próximos em comparação ao enfoque bayesiano com informação de parentesco e sem priori informativa. Utilizou-se um burn-in 15 e thin 6 iterações e a convergência foi alcançada.

Na Tabela 5 nota-se que as estimativas e predições de parâmetros de um modelo misto com enfoque Clássico usando o método REML/BLUP tiveram alguns resultados um pouco diferentes em comparação ao enfoque bayesiano com informação de parentesco com priori informativa. Utilizou-se um burn-in de 9 e thin de 3 iterações e a convergência foi alcançada.

Observando a Tabela 2 em comparação a 3, e a 4 em comparação a 5 verifica-se que a inclusão da distribuição *a priori* informativa faz com que todos os parâmetros diminuam. Quando se compara a Tabela 2 com a 4 e a Tabela 3

com a 5 verifica-se que a informação de parentesco ajuda a aumentar os efeitos genotípicos e a diminuir as variâncias e herdabilidade. Já na comparação da tabela 3 com a 4 verifica-se que informação de parentesco aumentou todos os parâmetros, o que torna evidente que a informação de parentesco tem um peso de informação maior em relação a priori informativa.

No entanto, não se pode desconsiderar que a priori informativa ajuda na diminuição das variâncias e herdabilidade e conforme Falconer & Mackay, 1996, a eficiência da seleção de famílias é preferida quando o caráter selecionado apresenta baixa herdabilidade. Contudo, a utilização dessas duas informações ajuda a aumentar os efeitos genotípicos e a diminuir as variâncias e herdabilidade como é observado na comparação da Tabela 2 com a 5.

Tabela 2 – Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (sem informação de parentesco e priori não informativa) para tonelada de colmos por hectare.

Parâmetros	Clássico		Bayesiano		
	Efeito	I.C. 95%		Efeito	I. Cred. 95%
$\sigma_g^2$	229,0589	109,75	740,15	301,5960	97,082 670,701
$\sigma_e^2$	538,5974	531,55	545,64	566,2199	415,964 769,401
$h^2$	0,6801	0,407	0,850	0,6958	0,135 0,562
$g_{61}$	26,3301	9,553	43,107	26,9133	7,311 47,338
$g_{65}$	17,3817	0,605	34,158	17,7771	-1,301 37,750
$g_{66}$	16,2544	-0,522	33,031	16,6632	-2,491 36,540
$g_{67}$	9,0838	-7,693	25,860	9,2933	-9,187 28,391
$g_{50}$	8,4911	-8,286	25,268	8,8697	-9,925 27,208
$g_{56}$	7,6944	-9,082	24,471	8,0567	-10,149 27,038
$g_{53}$	7,5002	-9,276	24,277	7,602	-10,620 26,936
$g_{54}$	6,7229	-10,054	23,500	6,9297	-11,007 25,424
$g_{45}$	3,3416	-13,435	20,118	3,3559	-14,617 21,969
$g_{49}$	2,1075	-14,669	18,884	2,0753	-16,336 20,732
$g_{55}$	1,3205	-15,456	18,097	1,4786	-16,832 20,075
$g_{68}$	-0,6131	-17,390	16,164	-0,6282	-19,209 18,013
$g_{60}$	-0,817	-17,594	15,960	-0,6957	-19,117 18,065
$g_{64}$	-2,9644	-19,741	13,812	-2,9524	-21,464 15,642
$g_{48}$	-5,3056	-22,082	11,471	-5,2977	-23,601 13,107
$g_{63}$	-5,4129	-22,190	11,364	-5,5547	-24,396 12,601
$g_{52}$	-8,182	-24,959	8,595	-8,4902	-26,789 9,682
$g_{51}$	-8,7649	-25,542	8,012	-8,9974	-27,994 9,296
$g_{46}$	-10,3391	-27,116	6,438	-10,5774	-29,700 8,015
$g_{57}$	-19,7541	36,531	-2,977	-20,2289	-40,385 -1,108
$g_{59}$	-22,0277	-38,804	-5,251	-22,6229	-42,809 -3,605
$g_{62}$	-22,0471	-38,824	-5,270	-22,4056	-42,230 -3,555

$\sigma_g^2$ : variância genotípica,  $\sigma_e^2$ : variância residual,  $h^2$ : herdabilidade da média da família,  $g_i$ : efeito genotípico da i-ésima família

Tabela 3 – Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (sem informação de parentesco e priori informativa) para tonelada de colmos por hectare.

Parâmetros	Clássico		Bayesiano	
	Efeito	I.C. 95%	Efeito	I. Cred. 95%
$\sigma_g^2$	229,0589	109,75 740,15	241,2471	142,472 390,060
$\sigma_e^2$	538,5974	531,55 545,64	544,4732	444,013 663,199
$h^2$	0,6801	0,407 0,850	0,6799	0,557 0,790
$g_{61}$	26,3301	9,553 43,107	26,3281	8,295 44,645
$g_{65}$	17,3817	0,605 34,158	17,47785	-0,511 35,592
$g_{66}$	16,2544	0,522 33,031	16,33948	-1,344 34,322
$g_{67}$	9,0838	-7,693 25,860	9,1121	-8,716 27,057
$g_{50}$	8,4911	-8,286 25,268	8,606767	-9,162 26,148
$g_{56}$	7,6944	-9,082 24,71	7,68656	-9,972 25,508
$g_{53}$	7,5002	-9,276 24,277	7,543723	-10,397 25,229
$g_{54}$	6,7229	-10,054 23,500	6,693979	-11,122 24,762
$g_{45}$	3,3416	-13,435 20,118	3,38317	-14,272 21,115
$g_{49}$	2,1075	-14,669 18,884	2,14487	-15,835 19,407
$g_{55}$	1,3205	-15,456 18,097	1,39622	-16,977 19,005
$g_{68}$	-0,6131	-17,390 16,164	-0,5798	-18,453 16,960
$g_{60}$	-0,817	-17,594 15,960	-0,88025	-18,531 16,967
$g_{64}$	-2,9644	-19,741 13,812	-2,97902	-20,735 14,641
$g_{48}$	-5,3056	-22,082 11,471	-5,36547	-23,267 12,485
$g_{63}$	-5,4129	-22,190 11,364	-5,29971	-23,472 11,903
$g_{52}$	-8,182	-24,959 8,595	-8,06953	-26,290 9,146
$g_{51}$	-8,7649	-25,542 8,012	-8,80558	-26,575 8,581
$g_{46}$	-10,3391	-27,116 6,438	-10,2793	-28,532 6,991
$g_{57}$	-19,7541	36,531 -2,977	-19,7063	-37,856 -1,929
$g_{59}$	-22,0277	-38,804 -5,251	-21,9477	-40,219 -4,539
$g_{62}$	-22,0471	-38,824 -5,270	-22,0414	-40,238 -4,373

$\sigma_g^2$ : variância genotípica,  $\sigma_e^2$ : variância residual,  $h^2$ : herdabilidade da média da família,  $g_i$ : efeito genotípico da i-ésima família

Tabela 4 – Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (com informação de parentesco e priori não informativa) para tonelada de colmos por hectare.

Parâmetros	Clássico			Bayesiano		
	Efeito	I.C. 95%		Efeito	I. Cred. 95%	
$\sigma_g^2$	229,0589	109,75	740,15	295,4213	98,278	669,842
$\sigma_e^2$	538,5974	531,55	545,64	567,2167	413,973	772,473
$h^2$	0,6801	0,407	0,850	0,6923	0,439	0,864
$g_{61}$	26,3301	9,553	43,107	27,9316	7,978	48,218
$g_{65}$	17,3817	0,605	34,158	18,2076	-0,682	37,747
$g_{66}$	16,2544	-0,522	33,031	17,0779	-1,853	36,585
$g_{67}$	9,0838	-7,693	25,860	9,6387	-8,965	28,923
$g_{50}$	8,4911	-8,286	25,268	10,3587	-8,453	29,586
$g_{56}$	7,6944	-9,082	24,471	8,1471	-10,236	27,25
$g_{53}$	7,5002	-9,276	24,277	6,7703	-12,591	25,691
$g_{54}$	6,7229	-10,054	23,500	9,7827	-8,686	28,813
$g_{45}$	3,3416	-13,435	20,118	3,1184	-14,988	22,448
$g_{49}$	2,1075	-14,669	17,884	0,9079	-17,511	19,547
$g_{55}$	1,3205	-15,456	18,097	1,7801	-16,801	20,366
$g_{68}$	-0,6131	-17,390	16,164	-0,1576	-19,124	18,256
$g_{60}$	-0,817	-17,594	15,960	-0,4726	-18,331	18,458
$g_{64}$	-2,9644	-19,741	13,812	-1,2729	-20,231	16,924
$g_{48}$	-5,3056	-22,082	11,471	-4,9298	-23,639	13,470
$g_{63}$	-5,4129	-22,190	11,364	-4,8907	-23,862	13,164
$g_{52}$	-8,182	-24,959	8,595	-5,2224	-24,347	13,271
$g_{51}$	-8,7649	-25,542	8,012	-8,5574	-27,474	10,030
$g_{46}$	-10,3391	-27,116	6,438	-10,7033	-29,712	7,902
$g_{57}$	-19,7541	36,531	-2,977	-19,641	-39,951	-0,659
$g_{59}$	-22,0277	-38,804	-5,251	-21,9287	-42,381	-2,885
$g_{62}$	-22,0471	-38,824	-5,0	-21,9449	-41,943	-2,912

$\sigma_g^2$ : variância genotípica,  $\sigma_e^2$ : variância residual,  $h^2$ : herdabilidade da média da família,  $g_i$ : efeito genotípico da  $i$ -ésima família

Tabela 5 – Estimativas, predições, intervalos de confiança (I.C.) e de credibilidade (I. Cred.) de parâmetros de um modelo misto com enfoque Clássico (REML/BLUP) e Bayesiano (com informação de parentesco e priori informativa) para tonelada de colmos por hectare.

Parâmetros	Clássico			Bayesiano		
	Efeito	I.C. 95%		Efeito	I. Cred. 95%	
$\sigma_g^2$	229,0589	109,75	740,15	240,1148	144,047	392,830
$\sigma_e^2$	538,5974	531,55	545,64	544,3562	444,888	665,155
$h^2$	0,6801	0,407	0,850	0,6793	0,555	0,792
$g_{61}$	26,3301	9,553	43,107	27,41696	9,220	46,363
$g_{65}$	17,3817	0,605	34,158	17,7736	-0,040	35,540
$g_{66}$	16,2544	-0,522	33,031	16,5408	-0,882	34,813
$g_{67}$	9,0838	-7,693	25,860	9,4663	-8,015	27,372
$g_{50}$	8,4911	-8,286	25,268	10,2149	-7,216	28,215
$g_{56}$	7,6944	-9,082	24,471	7,9132	-9,867	25,689
$g_{53}$	7,5002	-9,276	24,277	6,2672	12,106	-24,515
$g_{54}$	6,7229	-10,054	23,500	9,8734	-7,992	27,507
$g_{45}$	3,3416	-13,435	20,118	3,0269	-14,905	21,411
$g_{49}$	2,1075	-14,669	18,884	0,8619	-16,879	18,631
$g_{55}$	1,3205	-15,456	18,097	1,6571	-15,773	19,630
$g_{68}$	-0,6131	-17,390	16,164	-0,2792	-17,695	17,686
$g_{60}$	-0,817	-17,594	15,960	-0,6127	-18,263	17,2210
$g_{64}$	-2,9644	-19,741	13,812	-1,3511	-18,681	16,905
$g_{48}$	-5,3056	-22,082	11,471	-4,9838	-22,790	13,623
$g_{63}$	-5,4129	-22,190	11,364	-4,9585	-22,861	12,910
$g_{52}$	-8,182	-24,959	8,595	-4,8791	-23,424	13,579
$g_{51}$	-8,7649	-25,542	8,012	-8,4239	-26,329	9,249
$g_{46}$	-10,3391	-27,116	6,438	-10,5772	-28,595	7,909
$g_{57}$	-19,7541	36,531	-2,977	-19,3588	-37,753	-1,407
$g_{59}$	-22,0277	-38,804	-5,251	-21,6986	-40,074	-3,215
$g_{62}$	-22,0471	-38,824	-5,270	-21,5557	-39,910	-3,922

$\sigma_g^2$ : variância genotípica,  $\sigma_e^2$ : variância residual,  $h^2$ : herdabilidade da média da família,  $g_i$ : efeito genotípico da i-ésima família

As quatro diferentes maneiras de implementação do modelo misto adotadas para a variável TCH pelo método bayesiano obtiveram as mesmas três melhores famílias que no método clássico, que foram a 61, 65 e 66.

Os dois métodos bayesiano sem informação de parentesco (Tabelas 2 e 3) obtiveram a mesma ordem de classificação por melhor valor genotípico em comparação ao método REML/BLUP.

Já os dois métodos bayesiano com informação de parentesco (Tabelas 4 e 5) tiveram sua ordem de classificação familiar por melhor valor genotípico um pouco modificada em comparação ao método REML/BLUP. Considerando as Tabelas 4 e 5, a família 50 obteve efeito genotípico maior que a 67, a família 54 obteve efeito genotípico maior que a 53, a família 63 efeito genotípico maior que a 48 e a família 62 obteve efeito genotípico maior que a 59.

Os resultados a respeito dos efeitos genéticos da família, tanto pelo método bayesiano quanto pelo REML/BLUP foram consideravelmente parecidos quando se fez as análises com 5 repetições, resultado este já comentado por Resende (2002), que cita sobre a existência de uma teoria assintótica bayesiana de forma que com grandes amostras se conseguem resultados similares aos obtidos com o método REML. Mas também vale ressaltar que a partir de certo tamanho da amostra, aumentos adicionais no tamanho da amostra não trazem ganhos compensatórios na precisão.

Pela Tabela 6, verifica-se que o intervalo de confiança para os efeitos genotípicos preditos tem amplitude menor que o intervalo de credibilidade. Acredita-se que no fato de que no método clássico se faz predições onde estas tendem a uma única média do parâmetro de interesse, já no método bayesiano se faz predições em que variam as médias da priori a medida que se realiza as iterações, e para isso a cada iteração os valores da média da distribuição *a priori* oscilam muito até obter a convergência e assim alcançar uma distribuição *a posteriori* bem representativa.

Tabela 6 – Amplitude dos intervalos das estimativas e predições de parâmetros de um modelo misto com enfoque Clássico e Bayesiano para a variável TCH com número diferente de repetições.

Parâmetros	REML	Bayesiana			
		Sem IP Sem PI	Sem IP Com PI	Com IP Sem PI	Com IP Com PI
$\sigma_g^2$	630,4	573,6	247,6	571,6	248,8
$\sigma_e^2$	<b>14,1</b>	353,4	219,2	358,5	220,3
$h^2$	0,4	0,4	0,2	0,4	0,2
$g_{61}$	<b>33,6</b>	40,0	36,4	40,2	37,1
$g_{65}$	<b>33,6</b>	39,1	36,1	38,4	35,6
$g_{66}$	<b>33,6</b>	39,0	35,7	38,4	35,7
$g_{67}$	<b>33,6</b>	37,6	35,8	37,9	35,4
$g_{50}$	<b>33,6</b>	37,1	35,3	38,0	35,4
$g_{56}$	<b>33,6</b>	37,2	35,5	37,5	35,6
$g_{53}$	<b>33,6</b>	37,6	35,6	38,3	36,6
$g_{54}$	<b>33,6</b>	36,4	35,9	37,5	35,5
$g_{45}$	<b>33,6</b>	36,6	35,4	37,4	36,3
$g_{49}$	<b>33,6</b>	37,1	35,2	37,1	35,5
$g_{55}$	<b>33,6</b>	36,9	36,0	37,2	35,4
$g_{68}$	<b>33,6</b>	37,2	35,4	37,4	35,4
$g_{60}$	<b>33,6</b>	37,2	35,5	36,8	35,5
$g_{64}$	<b>33,6</b>	37,1	35,4	37,2	35,6
$g_{48}$	<b>33,6</b>	36,7	35,8	37,1	36,4
$g_{63}$	<b>33,6</b>	37,0	35,4	37,0	35,8
$g_{52}$	<b>33,6</b>	36,5	35,4	37,6	37,0
$g_{51}$	<b>33,6</b>	37,3	35,2	37,5	35,6
$g_{46}$	<b>33,6</b>	37,7	35,5	37,6	36,5
$g_{57}$	<b>33,6</b>	39,3	35,9	39,3	36,3
$g_{59}$	<b>33,6</b>	39,2	35,7	39,5	36,9
$g_{62}$	<b>33,6</b>	38,7	35,9	39,0	36,0

IP = informação de parentesco; PI= priori informativa,  $\sigma_g^2$ : variância genotípica,  $\sigma_e^2$ : variância residual,

$h^2$ : herdabilidade da média da família,  $g_i$ : efeito genotípico da i-ésima família

Por meio dos resultados da análise bayesiana mostrados na Tabela 6, observa-se que os intervalos de credibilidade para a variável TCH pelo método bayesiano sem informação de parentesco e com distribuição *a priori* informativa foram bem próximos dos resultados com informação de parentesco e com distribuição *a priori* informativa, e menores que os resultados obtidos pelo método sem informação de parentesco e sem distribuição *a priori* informativa que foi bem aproximado com os resultados com informação de parentesco sem distribuição *a priori* informativa.

No entanto, vale ressaltar que a matriz de parentesco obtida foi pouco informativa, como se pode observar na Tabela 7.

Tabela 7 – Matriz de Parentesco com 22 famílias

Famílias																						
	45	46	48	49	50	51	52	53	54	55	56	57	59	60	61	62	63	64	65	66	67	68
45	1	0	0	0	0	0.25	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0
46	0	1	0	0	0	0	0	0.25	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0
48	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	1	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0
50	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0
51	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	1	0	0.25	0	0	0	0	0	0.25	0	0	0	0	0	0.25	0
53	0	0	0	0	0	0	0	1	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0
54	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0.25	0	0	0	0	0	0.25	0
55	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
57	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0.25	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.25	0	0	0
65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
66	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Os resultados mostraram que a maior influência na obtenção de menores intervalos de credibilidade foi a utilização de uma distribuição *a priori* informativa, pois mesmo sem ou com a informação de parentesco os intervalos de confiança e credibilidade foram bem semelhantes. Entretanto, a informação de

parentesco fez com que ocorresse uma mudança na ordem de classificação das melhores famílias, mostrando que sua informação, por menor que seja, mudou de certo modo os resultados dos valores genotípicos.

Segundo Resende & Barbosa (2006) o método BLUPIS nos fornecem quantos indivíduos são selecionados em cada família utilizando informações apenas da família. Isto pode ser observado por meio da tabela 8 que teve como resultado o mesmo número de indivíduos a serem selecionados por família quando comparado o REML/BLUP com o bayesiano (Sem IP, Sem PI).

Através da tabela 8 se verifica que quanto maior o efeito, maior o número de indivíduos a ser selecionado, como também, a família com maior efeito genotípico é a 61 e com menor efeito é a 62. As famílias 68, 60, 64, 48, 63, 52, 51, 46, 57, 59 e 62 podem ser descartadas do estudo e se tiver a intenção de estudar o indivíduo dentro de cada família, analisa somente o número de indivíduos determinado das famílias que tiveram efeitos genotípicos positivos.

Contudo, acredita-se que o enfoque bayesiano em modelos mistos para o método BLUPIS se sobressai quando se utiliza a informação de parentesco e a distribuição *a priori* informativa juntas. Como Nogueira (2003) diz, a metodologia bayesiana se sobressai quando há informações conhecidas que são levadas em conta.

Tabela 8 - Valores dos efeitos genotípicos ( $\hat{g}_j$ ) e do número de indivíduos a serem selecionados ( $n_k$ ) em cada família utilizando a metodologia BLUP sob enfoque clássico e bayesiano para a característica tonelada de colmos por hectare (TCH) considerando o número de indivíduos a serem selecionados na melhor família ( $n_j$ ) igual a 50.

Famílias	$n_j = 50$					
	Clássico		bayesiano			
	$\hat{g}_j$	$n_k$	$\hat{g}_j$ (Sem IP Com PI)	$n_k$	$\hat{g}_j$ (Com IP Com PI)	$n_k$
$g_{61}$	26,3301	50	26,9133	50	27,4170	50
$g_{65}$	17,3817	33	17,7771	33	17,7736	32
$g_{66}$	16,2544	31	16,6632	31	16,5408	30
$g_{67}$	9,0838	17	9,2933	17	9,4663	17
$g_{50}$	8,4911	16	8,8697	16	10,2149	19
$g_{56}$	7,6944	15	8,0567	15	7,9132	14
$g_{53}$	7,5002	14	7,602	14	6,2672	11
$g_{54}$	6,7229	13	6,9297	13	9,8734	18
$g_{45}$	3,3416	6	3,3559	6	3,0269	6
$g_{49}$	2,1075	4	2,0753	4	0,8619	2
$g_{55}$	1,3205	3	1,4786	3	1,6571	3
$g_{68}$	-0,6131	0	-0,6282	0	-0,2792	0
$g_{60}$	-0,817	0	-0,6957	0	-0,6127	0
$g_{64}$	-2,9644	0	-2,9524	0	-1,3511	0
$g_{48}$	-5,3056	0	-5,2977	0	-4,9838	0
$g_{63}$	-5,4129	0	-5,5547	0	-4,9585	0
$g_{52}$	-8,182	0	-8,4902	0	-4,8791	0
$g_{51}$	-8,7649	0	-8,9974	0	-8,4239	0
$g_{46}$	-10,3391	0	-10,5774	0	-10,5772	0
$g_{57}$	-19,7541	0	-20,2289	0	-19,3588	0
$g_{59}$	-22,0277	0	-22,6229	0	-21,6986	0
$g_{62}$	-22,0471	0	-22,4056	0	-21,5557	0

#### 4 CONCLUSÕES

O método BLUPIS com enfoque bayesiano realizado através do algoritmo construído junto ao programa R foi muito eficiente e de grande ajuda para um melhor desempenho e rapidez na aplicabilidade de todos os cálculos.

O método bayesiano se mostrou mais eficiente, isto é, com efeitos genotípicos maiores e variâncias e herdabilidade menores, quando se consideraram no modelo a informação de parentesco e a distribuição da priori informativa.

Os resultados deste estudo não encontraram intervalos de credibilidade menores que os de confiança, mas estes não foram muito discrepantes.

#### 5 REFERÊNCIAS

BARBIN, D. **Componentes de variância: teoria e aplicações**. 2.ed. Piracicaba: FEALQ. 1993. p39-48

GEWEKE, J. Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. **Bayesian Statistics**. Oxford University Press, New York. v.4, 1992.

GIANOLA, D. **Bayesian analysis with applications to genetics e biology**. Ribeirão Preto: USP, 1996. Class notes of a course taught at the departamento of genetics, Medical School, University of São Paulo at Ribeirão Preto, 1996.

HENDERSON, C. R. et al. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**. v.15. 1959, p.192.

HENDERSON, C. R. Sire evolution and genetic trends. In: Animal Breeding and Genetics Symposium. 10.. 1973. Champaign. **Proceedings...** Champaign: American Society of Animal Science. 1973, p.10-41.

KIMBENG, C.A.; COX. M.C. Early generation selection of sugarcane families and clones in australia: a review. **Journal American Society of sugarcane Technologists**. v.23, 2003, p.20-39.

LACHOS, V.H. **Modelos Lineares Mistos Assimétricos**. Tese de Doutorado.IME-USP, 2005.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. **Melhoramento da cana-de-açúcar**. In: Borém, A. (2ª ed.). Melhoramento de espécies cultivadas. UFV: Viçosa. 2005, p.225-274.

NOGUEIRA, D. A. et al. Análise clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Cienc. Agrotec.**, Lavras Edição Especial, dez., 2003, p. 1614-1624.

RAFTERY, A.L.; LEWIS, S. chapter How Many Iterations in the Gibbs Sampler? **Bayesian Statistics**. Oxford University Press, New York. v.4, 1992.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing. Vienna, Áustria, 2011. Disponível:<<http://www.R-project.org>>.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica. 2002a, p.975.

RESENDE, M.D.V. **Software Selegen-Reml/Blup**. Curitiba: Embrapa Floresta. 2002b, 67p (Documento 77).

RESENDE, M. D. V. de; BARBOSA M.E.P. Selection via simulated individual BLUP based on family genotypic effects in sugarcane. **Pesq. agropec. bras.**. Brasília. v.41. n.3. mar. 2006, p.421-429.

SORENSEN, D. **Gibbs sampling in quantitative genetics**. Foulun:[s/n], 1996, p.186.

SORENSEN, D; GIANOLA, D. **Likelihood, Bayesian and MCMC Methods in Quantitative Genetics**. Springer. 2002. p.760.

SMITH, B. J. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. **Journal of Statistical Software**: Issue 11. v.21, 2007, p.1-41.

## CAPÍTULO II – DESENVOLVIMENTO DAS FUNÇÕES NO R

### 1 INTRODUÇÃO

Houve um crescente avanço da metodologia Bayesiana em genética. Ainda assim, existem relativamente poucas publicações que realizam essa metodologia, pois a maior parte das análises estatísticas tem certa dificuldade na resolução por forma analítica. Recorrer ao método computacional é mais fácil e prático na aplicação de problemas complexos, entretanto muitos softwares têm suas restrições e limites de uso.

É de grande ganho a disponibilização de programas computacionais que permitem ao usuário montar e fazer suas próprias rotinas. Isso torna possível realizar várias metodologias que ainda não são encontradas em software, como pode ser visto pelo artigo de Waldmann (2009) que realiza um código no programa WinBugs (*The BUGS Project*) para um modelo genético aditivo poligênica individual. Como também por Peternelli (2009), em que é implementado junto ao programa R uma rotina capaz de executar a análise de parentesco para qualquer nível 2k de ploidia.

Um dos softwares livre com uma linguagem dirigida à análise gráfica e estatística é o Programa R (*The R Project for Statistical Computing*). As vantagens desse programa é que além dos pacotes, ele fornece muitas funções já prontas, como também permite que os usuários adicionem novas funções e programem novas rotinas, ajudando dessa forma na realização de todo o processo de um algoritmo em que se deseja construir.

O programa pode ser obtido através do site <http://www.R-project.org>, fazendo o Download R 2.13. Ao instalar o R, apenas alguns pacotes vêm juntos, os quais são essenciais para o funcionamento do programa. Os pacotes extras podem ser encontrados com base no próprio programa, clicando em “pacotes” no menu, depois em “instalar pacote”, assim abrirá uma janela em que seleciona o nome do pacote desejado e depois é só clicar em “OK” e logo aparecerá no console do R um texto indicando que o pacote selecionado foi instalado.

Dúvidas a respeito das funções do programa podem ser tiradas por meio do livro de Peternelli e Mello (2011).

Verifica-se na literatura que há vários estudos relacionados ao melhoramento genético. O método de melhor predição linear não viesado (BLUP) (*Best Linear Unbiased Predictor*) individual simulado (BLUPIS) via enfoque bayesiano é um deles, que é utilizado quando se deseja selecionar famílias superiores.

A análise bayesiana realiza com segurança as inferências com parâmetros genéticos, valores genéticos e efeito de grupos genéticos (RESENDE, 2002), mas sua realização não é fácil, pois a obtenção de distribuições marginais por processos analíticos é praticamente impossível (GIANOLA, 1996).

Pensando nisso, o objetivo desse trabalho foi o de desenvolver e apresentar um algoritmo para o usuário em geral sob a forma de conjunto de funções de fácil utilização que realiza o método BLUPIS via enfoque bayesiano aplicado no delineamento mais utilizado de todos na experimentação; o Delineamento em Blocos Casualizados (DBC) (PETERNELLI; RESENDE; MENDES, 2012).

## **2 MATERIAL E MÉTODOS**

Todas as funções desenvolvidas foram escritas em linguagem R de programação em um computador com sistema operacional Windows. Foi montada uma sequência finita de ações que realizam todo o processo de estimação e predição de valores genéticos através de modelos mistos com enfoque bayesiano utilizando-se do método amostragem de gibbs.

Os pacotes utilizados para a realização do algoritmo desenvolvido neste trabalho foram os seguintes: um pacote que realiza a inversa da matriz de parentesco, um pacote necessário para utilizar a função *rinvchisq* e um pacote que realiza a análise de convergência, respectivamente denominados, “MCMCpack” (MARTIN; QUINN; PARK, 2011), “geoR” (RIBEIRO JUNIOR; DIGGLE, 2001) e “boa” (SMITH, 2007) .

## Algoritmo

O algoritmo é dividido em 2 etapas,

Etapa 1:

Primeiramente faz-se a leitura dos dados em formato de tabela salva em txt, posteriormente deve-se informar o número de efeitos fixos e o número de efeitos aleatórios, logo em seguida, entra-se com a matriz de parentesco. Com essas informações o algoritmo faz a álgebra de matrizes necessárias para a realização do modelo linear misto;

Etapa 2:

Começa-se a inferência bayesiana. Entra-se com os hiperparâmetros que auxiliam na especificação da distribuição *a priori* dos componentes de variância  $\sigma_g^2$  (variância genética aditiva) e  $\sigma_\varepsilon^2$  (variância residual), no caso desse trabalho, foi considerado uma distribuição qui-quadrado escalonada invertida, da forma:

$$p(\sigma_i^2 | v_i, S_i^2) \propto (\sigma_i^2)^{-\left(\frac{v_i}{2}+1\right)} \exp\left\{\frac{-v_i S_i^2}{2 \sigma_i^2}\right\}, \quad (i = g, \varepsilon)$$

em que  $v$  são os graus de liberdade da distribuição qui-quadrado e  $S_i^2$ , o valor inicial da variância.

Quando esta distribuição foi considerada não informativa foi substituído  $S_i^2$  por 0 e  $v$  por -2.

Posteriormente entra-se com os valores que inicializam a amostragem de gibbs, que são: o número de iterações que se deseja realizar, a variância genética aditiva e a variância do erro aleatório.

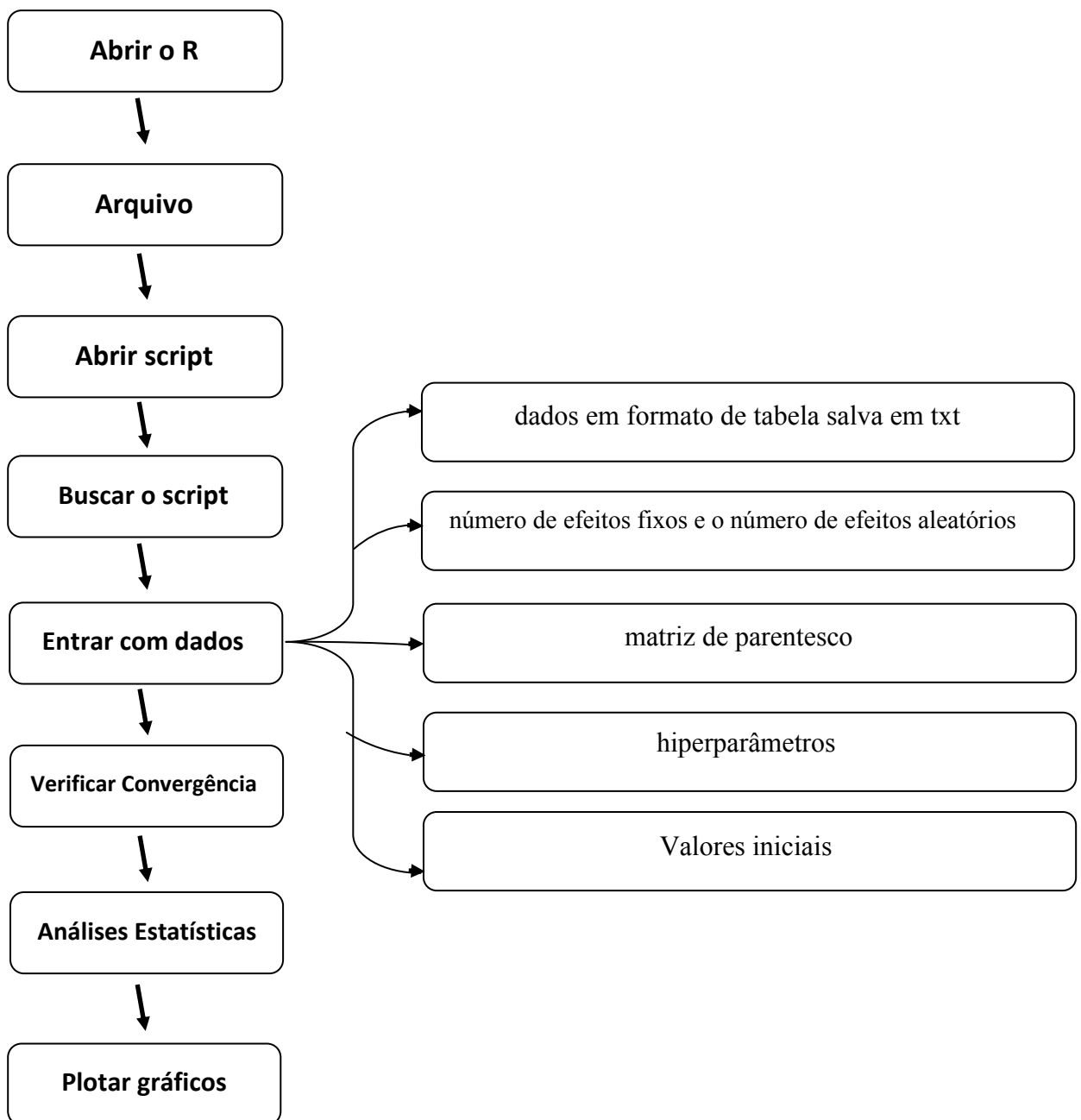
### 2.1 Procedimentos para a realização do algoritmo

A rotina desenvolvida pelo algoritmo foi exemplificada por meio de dados fictícios e baseado no modelo conforme apresentado no capítulo I.

O algoritmo foi adaptado para os experimentos com delineamento em blocos casualizados. O fluxograma do algoritmo e toda a sequência das funções é

dada a seguir, exemplificada em um DBC com 3 tratamentos (famílias) e 2 blocos (repetições).

**Figura 1 - Fluxograma do algoritmo**



```

> Este é um Algoritmo para o usuário em geral sob a forma de
conjunto de funções de fácil utilização que realiza o método
BLUPIS via enfoque bayesiano.
> aplicado no Delineamento em Blocos Casualizados(DBC).
> rm(list=ls(all=TRUE))
> #####----- PASSO A PASSO -----
#####
> # Modelo do exemplo:  $Y \sim Xb + Zg + e$  , com p blocos e q
tratamentos
> #####-----
#####
> ### Dados do Experimento ###

```

Para a realização do algoritmo, a primeira informação que o pesquisador deve dar é a respeito da identificação de cada família (trat), do bloco em que cada família se encontra (bloco) e o resultado da variável de interesse para cada família dentro de cada bloco (y). Estes dados devem ser colocados da maneira exposta abaixo, com estes mesmos nomes e salvos em arquivo txt (bloco de notas);

```

> dados<-read.table("dados.exemplo.txt",header =T)
> dados

```

Id	trat	bloco	y
1	1	1	150
2	1	2	143.75
3	2	1	180.35
4	2	2	198.21
5	3	1	160.71
6	3	2	166.07

A segunda informação é a do número de efeitos fixos (blocos) e o número de efeitos aleatórios (tratamentos);

```

> # _____ #
> p=2          ### número de efeitos fixos (blocos)
> q=3          ### número de efeitos aleatórios (tratamentos)
> # _____ #

```

```
> ### Matriz de parentesco genético aditivo ###
```

A matriz de parentesco é usada para dar informação do quão os genótipos são relacionados, sendo construída com base no coeficiente da parentesco de Malecot (coeficientes das covariâncias entre os efeitos aleatórios genéticos);

```
> ### Matriz de parentesco = Identidade ###
```

Se esta for igual à matriz de identidade, indica que o BLUP dos genótipos são preditos somente com base em suas próprias performances;

```
> A = diag(1,q,q)   ### matriz identidade
```

```
> AInv = solve(A)   ### solve é código do R para se fazer a inversa da
```

Matriz

```
> AInv
```

```
      [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

```
## Matriz de parentesco ###
```

Se o BLUP dos genótipos não é predito somente com base em suas próprias performances, então a matriz é montada através dos coeficientes de Malécot, que para sua construção necessita-se da informação dos pais dos indivíduos. Todo esse processo está bem explicado por PETERNELLI, L.A. et al. (2009).

Após a obtenção da matriz de coeficientes, salva-a em arquivo txt e dar-se um nome para ela, no exemplo exposto foi nomeada *matrizAdiv2*;

```
> matrizA <- read.table("matrizAdiv2.txt",header =F)
```

```
> A<-2*matrizA
```

```
> AInv = solve(A)   ### matriz inversa
```

```
> ### Matriz de Incidência dos blocos ###
```

Como já tem a informação dos dados, a coluna de blocos é selecionada com o comando abaixo e X é a matriz de incidência dos blocos;

```
> bloco<-dados$bloco
> d<-data.frame(blocof=as.factor(bloco))
> X<-model.matrix(~ blocof -1,d)
```

```
> X
      blocof1 blocof2
1           1         0
2           0         1
3           1         0
4           0         1
5           1         0
6           0         1
```

```
> XLX=t(X)%*%X    ### %*% Multiplicação de matrizes
```

**XLX é a multiplicação da matriz X transposta (X')** pela matriz X. Nota-se que o resultado é uma matriz diagonal do número de tratamentos em cada bloco;

```
> XLX
      blocof1 blocof2
blocof1         3         0
blocof2         0         3
```

```
> ### Matrizes de Incidência dos tratamentos ###
```

Como já tem a informação dos dados, a coluna de tratamentos é selecionada com o comando abaixo e Z é a matriz de incidência dos tratamentos;

```
> tratZ<-dados$trat
> d<-data.frame(trat=as.factor(tratZ))
> Z<-model.matrix(~ trat -1,d)
```

```
> Z
      trat1 trat2 trat3
1         1         0         0
2         1         0         0
3         0         1         0
4         0         1         0
5         0         0         1
6         0         0         1
```

```
> XLZ=t(X)%*%Z    ### %*% Multiplicação de matrizes
```

$XLZ$  é a multiplicação da matriz  $X$  transposta ( $X'$ ) pela matriz  $Z$ . Nota-se que o resultado é uma matriz que identifica quais blocos se encontram em cada tratamento;

```
> XLZ
      trat1 trat2 trat3
blocof1    1    1    1
blocof2    1    1    1
```

```
> W<- cbind(X, Z)
```

$W$  é matriz de incidência dos blocos e tratamentos juntos;

```
> W
      blocof1 blocof2 trat1 trat2 trat3
1          1          0      1      0      0
2          0          1      1      0      0
3          1          0      0      1      0
4          0          1      0      1      0
5          1          0      0      0      1
6          0          1      0      0      1
```

```
> WLW<- t(W) %*% W
```

$WLW$  é a multiplicação da matriz  $W$  transposta ( $W'$ ) pela matriz  $W$ ;

```
> WLW
      blocof1 blocof2 trat1 trat2 trat3
blocof1      3          0      1      1      1
blocof2      0          3      1      1      1
trat1        1          1      2      0      0
trat2        1          1      0      2      0
trat3        1          1      0      0      2
```

```
> ### vetor de resultados ###
```

```
> Y<-dados$y
```

$Y$  é o vetor dos resultados do experimento;

```

> Y
[1] 150.00 143.75 180.35 198.21 160.71 166.07
> n=length(Y)   ### tamanho do vetor Y

> n
[1] 6

> XLY=t(X)%*%Y

```

XLY é a multiplicação da matriz X transposta ( $X'$ ) pela matriz Y. Nota-se que o resultado é a soma dos tratamentos em cada bloco;

```

> XLY
      [,1]
blocof1 491.06
blocof2 508.03

```

```

> ZLY=t(Z)%*%Y

```

ZLY é a multiplicação da matriz Z transposta ( $Z'$ ) pela matriz Y. Nota-se que o resultado é a soma dos blocos em cada tratamento;

```

> ZLY
      [,1]
trat1 293.75
trat2 378.56
trat3 326.78

```

```

> WLY=t(W)%*%Y

```

WLY é a multiplicação da matriz W transposta ( $W'$ ) pela matriz Y. Nota-se que o resultado é soma dos tratamentos em cada bloco e a soma dos blocos em cada tratamento;

```

> WLY
      [,1]
blocof1 491.06
blocof2 508.03
trat1 293.75
trat2 378.56
trat3 326.78

```

```

> rj=c(WLY)   ### vetor de Média dos blocos e tratamentos

> rj
[1] 491.06  508.03  293.75  378.56  326.78

>
#####
#####
> #####----- INFERÊNCIA BAYESIANA -----
-----#####
>
#####
#####

> ### Pacotes ###
> library(geoR)           # Necessário para utilizar a função
rinvchisq
> library(MCMCpack)      # pacote para realizar as matrizes inversa
> library(boa)           # pacote para realizar a análise de
convergência

```

```

> ### Hiperparâmetros ###

```

A terceira informação que o pesquisador deve dar é a respeito dos hiperparâmetros, que são os parâmetros da distribuição *a priori*. A distribuição *a priori* dos componentes de variância considerada neste algoritmo foi a qui-quadrado escalonada invertida, onde  $\nu$  são os graus de liberdade e  $s^2$  o valor inicial da variância;

```

> ### Hiperparâmetros - Distribuição Uniforme Imprópria ###

```

A distribuição qui-quadrado escalonada invertida se reduz a uma distribuição uniforme imprópria se a os graus de liberdade são iguais a -2 e o valor inicial da variância é 0, e é utilizada quando se especifica uma noção de conhecimento vago, ou seja, segue uma distribuição constante que é conhecida como distribuição pouco informativa;

```

> S2e = 0; ve = -2   ### valor inicial da variância e g.l. dos
erros
> S2a = 0; va = -2   ### valor inicial da variância e g.l. dos
valores
genéticos aditivos

```

```

> ### Hiperparâmetros-Distribuição Qui-quadrado Escalonada
Invertida ###

```

A distribuição qui-quadrado escalonada invertida se torna própria quando se especifica valores para os hiperparâmetros diferentes dos dados anteriormente;

```

> S2e = 5; ve = 2   ### valor inicial da variância e g.l. dos
erros
> S2a = 5; va = 2   ### valor inicial da variância e g.l. dos
valores
genéticos aditivos

```

```

> ### Valores Iniciais ###

```

A quarta informação que o pesquisador deve dar é a respeito dos valores das variâncias que irão inicializar a amostragem de gibbs e o número de iteração que deseja realizar;

```

> # _____ #
> Namostra=30   ### número de iterações
> sigma2a=5     ### variância aditiva
> sigma2e=5     ### variância do erro aleatório
> # _____ #

```

```

> #####
> #####   Amostrador de Gibbs   #####
> #####

```

Agora, com todas as informações necessárias já obtidas, realiza-se o algoritmo do amostrador de gibbs que tem o objetivo de fazer várias iterações até a distribuição *a posteriori* convergir afim de obter bons resultados de interesse, que são: distribuições marginais a posteriori dos parâmetros genéticos, efeitos genotípico da família e valores genéticos;

```

> b<-numeric(p+q)
> tj<-numeric(p+q)
> t1<-numeric(p+q)
> t2<-numeric(p+q)
> k<-numeric(Namostra)
> amostra = matrix(0,ncol=(p+q+3),nrow=Namostra)
> ### p= número de efeitos fixos, q= número de efeitos aleatórios

> gibbs = function(Namostra){
>     amostra[1,] = cbind(t(b),sigma2a,sigma2e,0)
>     ### t(b) = transposta do vetor b que inicialmente é dada por
zeros
>     for(i in 2:Namostra){                                     # início do
for i
>         k[i] =as.numeric(amostra[(i-1),(p+q+2)]/amostra[(i-
1),(p+q+1)]) > #k=sigma2e/sigma2a = alfa
>         sigma = matrix(0,ncol=(p+q),nrow=(p+q))
>         sigma[(p+1):(p+q),(p+1):(p+q)] = AInv*k[i]
>     ### A matriz sigma na posição dos efeitos aleatórios será
representada     por Ainv*k
>         CC = WLW+sigma
>     ### CC é a matriz dos coeficientes das Eq. Mod. Mistos, WLW=W'W,
W=[X Z]

>         for(j in 1:(p+q)){                                     # início do
for j
>             CCjj = CC[j,j]                                     ### valor da linha j,
coluna j
>             CCj = as.vector(CC[j,])                           ### vetor da
linha j

>             if(j==1){
>                 tj = as.vector(amostra[(i-1),1:(p+q)])
>             }else{
>                 t1 = as.vector(b[1:(j-1)])
>                 t2 = as.vector(amostra[(i-1),j:(p+q)])
>                 tj = as.vector(c(t1,t2))
>             }
>             mj = (rj[j]-CCj[-j]*tj[-j])/CCjj

```

```

>           s2j = amostra[(i-1),(p+q+2)]/CCjj
>           b[j]= rnorm(1,mj,sqrt(s2j))
> #Distribuição condicional a posteriori dos blocos e tratamentos
> }                                     #fim do for j
>           amostra[i,1:(p+q)] <- b[]
>           a = b[(p+1):(p+q)]          ### só os efeitos
aleatórios
>           Va = q + va
>           S2A = (t(a)%*%AInv%*%a+(va*S2a))/Va
>           amostra[i,(p+q+1)] = rinvchisq(1,df=Va,scale=S2A)
> ### Sigma2a segue uma Distribuição Qui-Quadrado Escalonada
Invertida
>           Ve = n + ve
>           S2E = (t(Y-W%*%b)%*%(Y-W%*%b)+(ve*S2e))/Ve
>           amostra[i,(p+q+2)] = rinvchisq(1,df=Ve,scale=S2E)
> ### Sigma2e segue uma Distribuição Qui-Quadrado Escalonada
Invertida
amostra[i,(p+q+3)]=amostra[i,(p+q+1)]/(amostra[i,(p+q+1)]+(1/p)*amostra[i
,(p+q+2)])
> ### Herdabilidade
> }                                     #fim do
for i
colnames(amostra)=c("bloco1","bloco2","trat1","trat2","trat3","sig
ma2a","sigma2e","herdabilidade")
> return(amostra)
> }
> saida<-gibbs(Namostra)

```

Em seguida verificar quantas iterações iniciais serão excluídas (burn-in) da saída e a distância mínima entre uma iteração e outra para se obter uma subamostra independente (thin) indicado pelo método de convergência de Raftery e Lewis

```

> ### Análise de Convergência ###
> install.packages("boa")
> library("boa")
> boa.menu()
> 1 #1: File

```

```

> 3 #3: Import Data
> 5 #5: Data matrix Objects
> saida # Nome do objeto de interesse
> # se -> +++ Object successfully imported +++
> 1 #1: Back
> 1 #1: Back
> 3 #3: Analysis
> 4 #4: Convergence Diagnostics
> 6 #6: Raftery e Lewis

```

Após o resultado do método de convergência de Raftery e Lewis, dar-se outro nome para a saída com o burn-in e thin encontrados. No exemplo abaixo o nome dado foi `ss`.

```

> burn.in<-m      # burn-in encontrado pelo método de Raftery e
Lewis
> thin<-n        # thin encontrado pelo método de Raftery e Lewis
> burn.thin<-seq(1,nrow(saida),thin)
> s1<-saida[burn.thin,]
> ss<-s1[burn.in:length(s1[,1]),]

```

Em seguida verificar se a convergência foi alcançada pelo teste Geweke (1992).

```

boa.menu()
1 #1: File
3 #3: Import Data
5 #5: Data matrix Objects
ss # Nome do objeto de interesse
# se -> +++ Object successfully imported +++
1 #1:Back
1 #1:Back
3 #3: Analysis
4 #4: Convergence Diagnostics
4 #4: Geweke # Se p-valor for menor que 0,05, não existe evidência
de que convergiu.

```

A média, a densidade, o intervalo de credibilidade para cada parâmetro é encontrada da mesma maneira como posta abaixo, fazendo alteração apenas na representação da coluna.

```
> # Média da coluna 8, isto é, da herdabilidade.
> mean(ss[,8])
> # Função Densidade da variância aditiva
> plot(density(ss[,6]))
> #intervalo de credibilidade dos quantis da coluna 5, isto é, do
efeito genotípico do trat3
> quantile(ss[,5],prob=c(.025,.975))
```

O número de indivíduos a serem estudados em cada família selecionada é dado pela função abaixo. A família que tiver resultado negativo não terá nenhum indivíduo a ser analisado, i.é, esta família pode ser excluída do estudo.

```
> g<-numeric(q)
> nj=50      # número máximo de indivíduos a serem selecionados por
família
> for(k in (p+1):(p+q)){
> g[k]=mean(ss[,k])
> maxg=max(apply(ss[, (p+1):(p+q)],2,mean))# maior efeito genético
> n[k]=(g[k]/maxg)*nj  # número de indivíduos a serem
selecionados na família da posição k (Ex: família45 -> k=6)
> names(n)=c("bloco1","bloco2","trat1","trat2","trat3")
> }
> n[(p+1):(p+q)]
```

### 3 RESULTADOS

O conclusão final de todo o algoritmo é feita por meio da nova saída, denominada ss, onde se encontram todos os valores dos efeitos dos blocos (2 primeiras colunas), efeitos genotípicos (colunas trat1, trat2, trat3), variância aditiva (coluna sigma2a), variância do erro aleatório (coluna sigma2e), e herdabilidade (coluna herdabilidade).

A obtenção dos valores das iterações de cada parâmetro (coluna) permitem encontrar a distribuições *a posteriori*, a análise descritiva completa e a construção de intervalos de credibilidade referentes a cada parâmetro, além de suas representações gráficas.

Para a visualização do leitor, foi realizado somente 30 iterações, mas é necessário um número bem maior de iterações, que é indicado por meio do teste de convergência de Raftery e Lewis.

```
> saida
      bloco1  bloco2   trat1   trat2   trat3   sigma2a   sigma2e  herdabilidade
[1,] 0.0000   0.0000   0.00000  0.00000  0.0000000 5.000000e+00  5.00000  0.0000000
[2,] 162.2691 168.9771 -10.62206 16.52896 -0.3806835 1.997840e+03 55.91191 0.9862000
[3,] 165.4022 170.0223 -22.31757 15.48907 -7.3811339 3.222554e+02 67.31700 0.9054308
[4,] 167.8364 167.9471 -16.30306 28.78460 -3.8513333 2.354043e+03 78.29408 0.9836423
[5,] 154.4775 166.0784 -22.20906 22.88707 5.9023117 1.951320e+03 62.16441 0.9843209
[6,] 161.3906 176.0522 -17.36291 16.73019 -7.2771938 4.642765e+02 142.70453 0.8667878
[7,] 177.3110 170.2181 -21.37939 22.60856 -17.1061035 1.797369e+03 42.54243 0.9883038
[8,] 166.1523 171.9827 -27.33447 19.09260 -3.0561752 2.545694e+04 86.02546 0.9983132
[9,] 170.4820 174.0392 -31.98167 25.04963 1.0060816 1.365636e+04 132.03874 0.9951889
[10,] 168.3379 168.4157 -30.41468 4.44244 -23.6866630 2.269447e+04 295.78177 0.9935256
[11,] 184.8237 180.2125 -41.48023 13.90998 -24.6593206 5.269250e+02 154.75261 0.8719574
[12,] 180.1487 190.1578 -32.82689 -0.06602 -17.6484085 2.943669e+02 47.50406 0.9253360
[13,] 181.5021 186.1888 -31.94069 5.35415 -26.4736107 8.764770e+02 37.63373 0.9789825
[14,] 183.1660 188.3516 -42.05796 -2.63441 -19.1195680 2.584200e+04 53.22794 0.9989712
[15,] 180.7961 194.3572 -46.91548 0.51226 -14.6776252 3.182653e+04 30.63876 0.9995189
[16,] 182.7763 189.2999 -39.18457 2.35923 -27.0219446 1.266600e+03 56.61569 0.9781391
[17,] 189.0571 195.7124 -48.89627 -7.43516 -21.5012638 1.683989e+04 56.20982 0.9983338
[18,] 191.5150 198.1048 -50.28654 -3.93289 -32.7343150 1.963785e+03 21.63657 0.9945213
[19,] 192.5427 197.4306 -48.80788 -4.95346 -33.6932638 1.040718e+09 37.27252 1.0000000
[20,] 193.4717 192.6754 -43.88547 0.65208 -29.9253228 8.526495e+05 39.26319 0.9999770
[21,] 183.3671 194.7309 -50.47604 -1.32887 -19.2985033 1.542640e+06 234.79450 0.9999239
[22,] 191.3128 187.4614 -38.53473 6.60891 -25.7210627 1.044817e+03 128.10641 0.9422356
[23,] 183.6735 185.0389 -43.07651 12.78627 -36.1674984 1.306181e+03 402.90176 0.8663793
[24,] 192.2584 196.2818 -53.51491 -5.01859 -15.5057590 5.288455e+06 235.47041 0.9999777
[25,] 198.5592 198.1701 -58.79014 -13.52915 -31.5436221 7.721028e+03 80.05677 0.9948424
[26,] 198.5645 204.5740 -58.55070 -16.36720 -28.1305605 2.981713e+04 679.60488 0.9887322
[27,] 187.8131 213.1605 -17.01540 3.05787 -99.9041298 3.572301e+03 3321.31748 0.6826540
[28,] 194.9554 291.5692 -90.15223 -46.74051 -54.3074134 4.737918e+05 2871.67753 0.9969786
[29,] 205.7197 259.9189 -110.44116 -18.07961 -111.4715325 4.753339e+04 544.89935 0.9943009
[30,] 235.0539 228.1063 -111.72045 -61.03011 -47.4311160 1.407058e+04 526.81866 0.9816235
```

A média, a função densidade e o intervalo de credibilidade para o efeito genotípico do tratamento 3, foi realizado por meio dos seguintes comandos abaixo:

```
> mean(saida[,5])
> [1] -25.75889

> plot(density(saida[,5]),main= "Função Densidade do tratamento 3
")
```

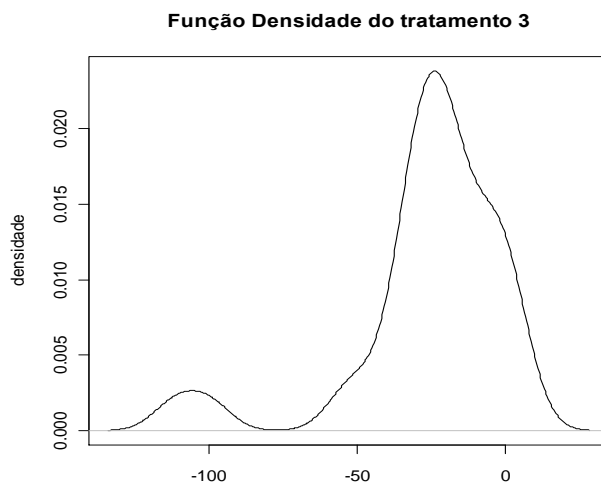


Figura 2 – Função Densidade do tratamento 3

```
> quantile(saida[,5],prob=c(.025,.975))
> 2.5%          97.5%
> -103.085166  2.352545
```

O número de indivíduos a serem estudados em cada família selecionada é obtida a seguir.

```
> g<-numeric(q)
> nj=50
> for(k in (p+1):(p+q)){
+ g[k]=mean(saida[,k])
+ maxg=max(apply(saida[, (p+1):(p+q)],2,mean))
+ n[k]=(g[k]/maxg)*nj
+ names(n)=c("bloco1","bloco2","trat1","trat2","trat3")
```

```

+ }
> n[(p+1):(p+q)]
      trat1      trat2      trat3
-1.813858   50.000000   6.308110

```

#### 4 CONCLUSÃO

O algoritmo foi de fácil aplicação e entendimento gerando os valores necessários para o estudo em análise.

O algoritmo foi testado em dados fictícios e em dados originais, sendo o resultado do segundo fornecido no capítulo 1.

A metodologia bayesiana permite uma análise descritiva completa para cada parâmetro.

#### 5 REFERÊNCIAS

GEWEKE, J. chapter Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. **Bayesian Statistics**. Oxford University Press, New York. v.4, 1992.

GIANOLA, D. **Bayesian analysis with applications to genetics e biology**. Ribeirão Preto: USP, 1996. Class notes of a course taught at the departamento of genetics, Medical School, University of São Paulo at Ribeirão Preto, 1996.

MARTIN, A. D.; QUINN K. M.; PARK J. H. MCMCpack: Markov Chain Monte Carlo in R. **Journal of Statistical Software**. 2011, 42(9), 1-21. URL <http://www.jstatsoft.org/v42/i09/>.

PETERNELLI, L.A.; FERREIRA, F.M.; ROCHA, R.B.; BARROS, W.S.; BARBOSA, M.H.P. Análise dos coeficientes de endogamia e de parentesco para qualquer nível de ploidia usando o pacote estatístico R. **Bragantia**. v.4, 2009.

PETERNELLI, L.A. ; MELLO, M.P. **Conhecendo o R: Uma visão estatística**. Viçosa, MG: Ed. UFV, 2011.

PETERNELLI, L.A.; RESENDE, M.D.V.; MENDES, T. O. P. . **Experimentação e análise estatística em cana-de-açúcar**. In: Fernando Santos; Aluizio Borém; Celso Caldas. (Org.). Cana-de-açúcar: Bioenergia, Açúcar e Etanol. 2 ed. Viçosa: Folha da Mata Ltda, 2012, v. 1, p. 337.

RAFTERY, A.L.; LEWIS, S. chapter How Many Iterations in the Gibbs Sampler? **Bayesian Statistics**. Oxford University Press, New York. v.4, 1992.

R DEVELOPMENT CORE TEAM, R: **A language and environment for statistical computing**. R Foundation for Statistical Computing. Vienna, Áustria, 2011, Disponível:<<http://www.R-project.org>>.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica. 2002, p.975.

RIBEIRO JUNIOR, P.J & DIGGLE, P. J. **geoR: a package for geostatistical analysis R-NEWS**, June, 2001, (2):15-18.

SMITH, B. J. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. **Journal of Statistical Software**, (2007) , 21(11), 1-37.

WALDMANN, P. Easy and flexible bayesian inference of quantitative genetic parametric. **Evolution**. 2009, 63-6:1640-1643.