



**ANÁLISE BAYESIANA DE UM MODELO  
LINEAR GENERALIZADO MISTO: EMPREGO  
NO MELHORAMENTO DE PLANTAS**

**JOSÉ WALDEMAR DA SILVA**

**2004**

57529  
049253

**JOSÉ WALDEMAR DA SILVA**

**ANÁLISE BAYESIANA DE UM MODELO LINEAR GENERALIZADO  
MISTO: EMPREGO NO MELHORAMENTO DE PLANTAS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária para a obtenção do título de “Mestre”.

Orientador



Prof. Dr. Júlio Sílvio de Sousa Bueno Filho

LAVRAS

MINAS GERAIS - BRASIL

2004

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

**Silva, José Waldemar da**

**Análise bayesiana de um modelo linear generalizado misto: emprego no melhoramento de plantas / José Waldemar da Silva. – Lavras : UFLA, 2004.**

**77 p. : il. .**

**Orientador: Júlio Sílvio de Sousa Bueno Filho.**

**Dissertação (Mestrado) – UFLA.**

**Bibliografia.**

**1. Componentes da variância. 2. Inferência bayesiana. 3. Modelos lineares generalizados mistos. 4. Modelo Poisson. I. Universidade Federal de Lavras. II. Título.**

**CDD-519.542**

**JOSÉ WALDEMAR DA SILVA**

**ANÁLISE BAYESIANA DE UM MODELO LINEAR GENERALIZADO  
MISTO: EMPREGO NO MELHORAMENTO DE PLANTAS**

Dissertação apresenta à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós-  
graduação em Agronomia, área de concentração em  
Estatística e Experimentação Agropecuária.

APROVADA em Terça-feira, 03 de fevereiro de 2004.

Prof. Dr. Daniel Furtado Ferreira  
Prof. Dr. Magno Antonio Patto Ramalho

UFLA  
UFLA



Prof. Dr. Júlio Sílvio de Sousa Bueno Filho

UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS – BRASIL  
2004

**A Deus,**

**Pela infinita bondade e amor e por me proporcionar a realização desta importante conquista.**

**OFEREÇO**

**Ao meu sempre e querido pai, Gregório Antônio da Silva**

**("in memorian"), pela coragem, incentivo e amor;**

**À minha maravilhosa mãe, Maria Edith Martins Silva,**

**pelos ensinamentos de vida, incentivo, amor e confiança;**

**À minha esposa, Elizângela, pela força, confiança,  
amizade e amor;**

**Aos meus irmãos, Valdete, José Valdo, Valter, Valdeir e Valdinei,**

**pela união, amizade e carinho;**

**DEDICO**

## AGRADECIMENTOS

A Deus, por ter me dado saúde, força e coragem para a realização deste trabalho.

À Universidade Federal de Lavras (UFLA) e ao programa de Pós-Graduação em Agronomia/Estatística e Experimentação Agropecuária pela oportunidade e confiança no meu trabalho.

À Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES) pela concessão da bolsa de estudos.

Ao professor orientador Júlio Sílvio de Sousa Bueno Filho pela orientação, sugestões e paciência na elaboração deste trabalho.

Aos professores do Departamento de Ciências Exatas, Augusto, Daniel, Delly, Eduardo, Joel, Júlio, Lucas, Luiz Henrique, Marcelo, Mário, Paulo e Thelma, pelos ensinamentos fundamentais para minha formação profissional e humana.

Ao Prof. Ednaldo Carvalho Guimarães pelo incentivo e por me iniciar na pesquisa científica.

Aos funcionários do Departamento de Ciências Exatas, Edila, Maria, Maristela, Meire, Selma e Sheila, pela eficiência que nos atenderam.

Aos colegas de curso Carla, Denis, Eliana, Fabyano, Francisco, Janser, Leite, Néia, Rafaela, Renata e Vânia pela amizade;

Ao colega Eduardo do Programa de Pós-Graduação em Agronomia/Genética e Melhoramento de Plantas pelo fornecimento dos dados para a exemplificação do trabalho.

Aos demais colegas que direta ou indiretamente estiveram presentes durante essa conquista.

## SUMÁRIO

	<b>Página</b>
RESUMO.....	i
ABSTRACT.....	ii
1 INTRODUÇÃO.....	01
2 REFERÊNCIAL TEÓRICO.....	03
2.1 Experimentação em melhoramento de plantas.....	03
2.2 Modelos lineares generalizados (MLG).....	05
2.2.1 Modelo Poisson.....	06
2.3 Modelos lineares generalizados mistos (MLGM).....	08
2.4 Inferência bayesiana.....	09
2.4.1 Priori não informativa.....	10
2.4.2 Procedimentos bayesianos para MLGM.....	11
2.4.3 Amostrador de Gibbs.....	13
2.4.4 Amostragem por rejeição com algoritmo Metropolis & Hastings.....	14
2.4.5 Diagnósticos de convergência.....	15
2.5 MLGM hierárquicos para dados Poisson.....	16
3 MATERIAL E MÉTODOS.....	20
3.1 Material.....	20
3.2 Metodologia.....	22
3.2.1 MLGM Poisson.....	22
3.2.2 Inferência bayesiana para MLGM Poisson.....	24
3.2.3 Posteriori conjunta.....	28
3.2.4 Forma de obtenção das condicionais completas.....	29
4 DESENVOLVIMENTO.....	30
4.1 Análise explotatória.....	30

4.2	Obtenção das condicionais completas.....	33
4.3	Características do Amostrador de Gibbs para algumas estatísticas de interesse.....	40
4.4	Implementação do processo amostral e diagnóstico de convergência.....	41
4.5	Estimativas de componentes da variância.....	42
4.6	Variância genética e herdabilidade.....	44
4.7	Estimativas e comparações entre efeitos médios.....	47
4.8	Distribuição dos efeitos médios de cruzamentos e de parcelas.....	50
5	CONCLUSÃO.....	53
	REFERÊNCIAS BIBLIOGRÁFICAS.....	54
	ANEXOS.....	57



## RESUMO

SILVA, José Waldemar da. *Análise bayesiana de um modelo linear generalizado misto: emprego no melhoramento de plantas*. 2004. 77 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.\*

A análise de dados experimentais frequentemente envolve tratamentos de efeitos aleatórios. São também relativamente frequentes dados de contagens para características de interesse, que em geral apresentam distribuição Poisson. Nestes casos a forma adequada de análise é ajustar modelos lineares generalizados mistos (MLGM). É comum a realização de inferências de forma assintótica usando aproximações normais do modelo descrito acima. Embora muitas vezes tais aproximações sejam satisfatórias, a inferência bayesiana permite a obtenção de distribuições exatas a posteriori para cada parâmetro. Neste trabalho foi aplicada a amostragem Gibbs para o ajuste de um MLGM usando a inferência bayesiana em um ensaio envolvendo dados de contagens de tubérculos graúdos em batata, visando a obtenção de estimativas de parâmetros genéticos como herdabilidades, componentes da variância e valores genéticos. Para ilustrar a metodologia foram utilizados dados experimentais do Programa de Melhoramento Genético de Batata da Universidade Federal de Lavras (UFLA). Implementou-se um algoritmo no software R para a amostragem Gibbs das distribuições a posteriori nas quais se pôde estimar os parâmetros genéticos e fenotípicos de interesse. Foram também tomadas amostras de combinações lineares e não lineares dos parâmetros originais do modelo, gerando novas distribuições a posteriori e respectivas inferências. Em virtude dos resultados apresentados, pode-se atestar que a inferência bayesiana é uma forma adequada e confiável para a análise de dados experimentais em genética.

**Palavras Chaves:** batata, componentes da variância, inferência bayesiana, modelos lineares generalizados mistos, modelo Poisson.

---

\* Orientador: Prof. Dr. Júlio Sílvio de Sousa Bueno Filho – UFLA.

## ABSTRACT

SILVA, José Waldemar da. **Bayesian annalysis of a generalized linear mixed model in plant breeding**. 2004. 77 p. Dissertation (Master in Statistics and Agricultural Experimentation) - Universidade Federal de Lavras, Lavras, MG.\*

Analysing experimental data in genetics is a common assumption that treatment effects comes from a probability distribution. Count data that shows Poisson distribution are also common responses for many economic important traits. In this case the correct analysis is based on fitting a generalized mixed linear model (GMLM). Assimptotic inference using normal approximation of the above model is a common choice. Although this usually results in good analysis, bayesian inference makes possible to get the exact posterior distributions for each parameter. In this work we had implemented Gibbs sampling to fit a GMLM using bayesian inference on data from a field trial on numbers of large tubers in potato, to get estimates of genetic parameters as heritabilities, variance components and genetic values. As an example it was used experimental data from the potato breeding program of the Universidade Federal de Lavras (UFLA). A Gibbs Sampling algorithm was implemented on R to get the posterior distributions on which inference on each genetic and phenotypic parameter can be done. Samples from linear and non-linear combinations of the original parameters of the model wer also taken, getting addictional posterior distributions and respective inferences. Results support the idea that bayesian inference is na elegant and reliable way to annalyse genetic experiments.

**Key Words:** bayesian inference, variance components, generalized mixed linear models, Poisson models, potato.

---

\* Adviser: Prof. Dr. Júlio Sílvio de Sousa Bueno Filho – UFLA.

## 1 INTRODUÇÃO

Nas diversas áreas da pesquisa agronômica o pesquisador depara-se com situações em que variável a ser analisada é contagem, e ainda é necessário atender às pressuposições de normalidade e homogeneidade da variância para a utilização do teste F. Em outras situações os modelos a serem ajustados envolvem efeitos fixos e aleatórios: no primeiro caso o interesse está nas médias dos fatores e, no segundo, o interesse são as estimativas das componentes da variância ou, nos parâmetros associados à variabilidade das variáveis respostas.

A teoria de modelos lineares generalizados (Nelder & Wedderburn, 1972) propiciou uma melhor tratabilidade de dados, pois admite-se, nesta teoria, qualquer distribuição da família exponencial. Assim, a única pressuposição que deve ser atendida para a análise com as técnicas de modelos lineares clássicos é a de independência. O objetivo desta metodologia é usar uma função que liga os valores observados aos valores esperados de uma combinação linear dos efeitos sistemáticos.

Na área de genética o uso da teoria de modelos lineares mistos tem se mostrado uma técnica de grande importância, visto que, nesta área, a análise de componentes da variância é de grande utilidade, permitindo cálculo de herdabilidades para a seleção de determinadas características.

Segundo Schall (1991), dois problemas são encontrados em modelos lineares generalizados mistos. O primeiro é a dificuldade para justificar uma distribuição para os efeitos aleatórios. O segundo é que a estimação baseada em máxima verossimilhança envolve integrais que em certos casos não têm soluções analíticas, tornando-se assim, necessário o uso de algoritmos de integração numérica.

A possibilidade de estabelecer uma distribuição para os parâmetros e através dela realizar inferências exatas, tem demonstrado uma grande motivação

para a utilização da inferência bayesiana. Isto porque, muitas vezes, problemas que têm solução complicada na análise frequentista têm solução bayesiana simples e direta.

Uma das vantagens de realizar inferências estatísticas por meio da teoria bayesiana sobre os parâmetros ou sobre dados não observados é que elas são feitas em termos de probabilidade estabelecida (Gelman et al., 1995).

A abordagem bayesiana possibilita a construção de intervalos de credibilidade para as estimativas dos parâmetros, estando a confiabilidade (precisão) dos parâmetros associada com a quantidade e precisão das informações utilizadas (Resende, 2002).

Tamanho amostral finito não é problema para a teoria bayesiana. Para qualquer conjunto de dados, pequeno ou grande, existirá uma distribuição a posteriori exata para a realização de inferências. Um outro fato a ser considerado em análise bayesiana é que as inferências realizadas são condicionadas aos dados, ou seja, permitem a análise completa sem violar o Princípio da Verossimilhança, no qual toda a informação contida na amostra ou experimento encontra-se representada na função de verossimilhança.

Assim, incorporando na análise todas as informações a priori e todas as que provem do experimento, obtém-se uma distribuição a posteriori conjunta, na qual devem ser realizadas as inferências de interesse.

Este trabalho teve por objetivo utilizar a metodologia de modelos lineares generalizados mistos e inferência bayesiana na análise de um experimento de melhoramento de batata em que o interesse foi a obtenção de estimativas de parâmetros genéticos como herdabilidades, componentes da variância e valores genéticos dos cruzamentos.

## 2 REFERENCIAL TEÓRICO

### 2.1 Experimentação em melhoramento de plantas

Em genética é comum considerar fatores de um modelo linear como aleatório. Por exemplo, considerando efeitos genéticos como fator aleatório, pode se estimar a variabilidade do genótipo para determinada característica. Se a variabilidade genética é alta, comparada com a ambiental, tem-se uma boa oportunidade para realizar seleção. Neste caso, os genótipos no experimento são considerados como uma amostra da população de genótipos na qual se deseja realizar inferências.

Quando for razoável assumir que uma amostra de indivíduos, como os níveis de um determinado fator, provém de uma população de níveis e que estes são representativos de tal população, então são considerados de efeitos aleatórios. Modelos que possuem efeitos fixos e aleatórios, podem ser escritos como segue:

$$y = X\beta + Zu + \varepsilon, \quad (2.1)$$

em que  $y$  é o vetor de observações de dimensões  $(n \times 1)$ ,  $n$  o número de observações,  $X$  a matriz de incidência do modelo para efeitos fixos  $(n \times p)$ ,  $p$  o número de parâmetros associados aos efeitos fixos,  $\beta$  o vetor de parâmetros fixos  $(p \times 1)$ ,  $Z$  a matriz de incidência para efeitos aleatórios  $(n \times q)$ ,  $q$  o número de parâmetros associados aos efeitos aleatórios,  $u$  o vetor de parâmetros aleatórios e  $\varepsilon \sim N(\emptyset, I\sigma^2)$  o vetor de resíduos  $(n \times 1)$ .

Nos modelos lineares, efeitos fixos são usados para modelar a média enquanto os efeitos aleatórios estabelecem a estrutura de variância-covariância

da variável resposta ( $y$ ). Uma das razões para a ocorrência de efeitos aleatórios é a simplificação na especificação dos distintos componentes da variância ( $\text{var}(y)$ ) (McCulloch & Searle, 2001), importantes parâmetros na pesquisa genética. Segundo os mesmos autores, a pressuposição de normalidade é requerida tanto para os modelos de efeitos fixos quanto para os modelos mistos.

A eficiência dos delineamentos do tipo blocos casualizados ou quadrados latinos está relacionada com o controle local efetuado pelos blocos ou colunas e linhas destes delineamentos. Porém, a quantidade de materiais a serem avaliados em melhoramento de plantas, em certos experimentos, é muito grande; desta forma, torna-se impossível obter áreas extensas e homogêneas de tal forma que o controle local seja satisfatório.

Para solucionar tal problema, Yates (1936) introduziu o delineamento em blocos incompletos, nos quais o número de parcela por blocos é menor que o número de tratamentos.

O delineamento em látice quadrado com tratamentos adicionais, um tipo de delineamento de blocos incompletos, pode ser utilizado quando se deseja uma precisão maior na comparação de materiais a serem melhorados com outros de comportamento já conhecido (Ramalho et al., 2000).

Muitos modelos na experimentação não apresentam erros normais devido ao fato de que as variáveis podem ser contagens e, portanto, provêm de uma distribuição Poisson. Nestes casos, os pesquisadores usam transformações (Bartlett, 1947) para atender esta pressuposição requerida pelos modelos lineares e modelos lineares mistos. Hinkelmann & Kempthorne (1994) apresentam algumas transformações de dados para atender as pressuposições de aditividade e homogeneidade da variância na análise de dados experimentais, tais como  $\sqrt{y}$  para dados de contagens e  $\arcsen\sqrt{y}$  para dados de proporções.

O objetivo da transformação é obter os dados em uma escala mais apropriada para que a análise seja viável, porém a escolha de uma transformação que reproduza todas as pressuposições simultaneamente não é tarefa fácil.

## 2.2 Modelos Lineares Generalizados

Para atender a pressuposição de normalidade para os resíduos em modelos lineares, geralmente usam-se transformações do tipo logarítmica, raiz quadrada e outras. Uma generalização dos modelos de distribuição normal são os modelos lineares generalizados (MLG), propostos por Nelder & Wedderburn (1972) em que o objetivo é analisar os dados de acordo com suas características, ou seja, utilizar uma distribuição adequada para a natureza das observações e aplicar uma transformação nos valores esperados de interesse (McCulloch & Searle, 2001), chamada função de ligação.

O modelo deve ser estruturado definindo três componentes:

1 – Uma distribuição para as variáveis aleatórias. As observações em geral são assumidas como medidas independentes e pertencentes a uma distribuição da família exponencial

$$y_i \sim \text{indep. } f_{y_i}(y_i)$$
$$f_{y_i}(y_i) = \exp \left\{ \frac{1}{\phi} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\} \quad (2.2)$$

em que  $\theta_i$  e  $\phi$  são parâmetros e  $b(\theta_i)$  e  $c(y_i, \phi)$ , funções conhecidas.

2 – Variáveis preditoras ou explicativas que são expressas como uma combinação linear dos efeitos.

$$\eta = X\beta,$$

em que  $\eta$  é o vetor de preditores lineares ( $n \times 1$ ).

3 – Uma função de ligação entre a média e uma combinação linear das variáveis preditoras.

$$\begin{aligned} E[y_i] &= \mu_i \\ \eta_i = g(\mu_i) = x_i' \beta &\Rightarrow \mu_i = g^{-1}(\eta_i), \end{aligned} \quad (2.3)$$

em que  $g(\cdot)$  é uma função conhecida, chamada função ligadora.

Existem algumas funções candidatas naturais a ligações para algumas distribuições de probabilidade mais comuns. Estas são as chamadas funções de ligação canônicas, observadas ao colocar a distribuição de probabilidade na forma exponencial.

### 2.2.1 Modelo Poisson

Vários fenômenos observáveis na experimentação apresentam dados de contagens; assim a distribuição Poisson se torna de grande utilidade e aplicação.

Segundo Mood et al. (1974), uma variável aleatória  $Y$  é definida como tendo uma distribuição Poisson se a densidade de  $Y$  é dada por

$$f_Y(y) = \frac{e^{-\mu} \mu^y}{y!} I_{\{0,1,2,\dots\}}(y), \quad \mu > 0, \quad (2.4)$$



a qual pode ser escrita como segue:

$$f_{y_i}(y_i) = \exp \{y_i \log(\mu_i) - \mu_i - \log(y_i!)\}. \quad (2.5)$$

Nota-se que (2.5) é uma forma exponencial como em (2.2) com as seguintes particularidades:

$$\phi = 1, \theta_i = \log(\mu_i), b_i(\theta_i) = \mu_i \text{ e } c_i(y_i, \phi) = -\log(y_i!).$$

Assim, verifica-se que uma função de ligação canônica para o caso Poisson é

$$\eta_i = g(\mu_i) = \log(\mu_i).$$

A Tabela 1 mostra algumas funções ligadoras para suas respectivas distribuições.

**TABELA 1: Distribuições de probabilidades com suas respectivas funções de ligação canônicas.**

Distribuições de probabilidade	Função de ligação canônica
Normal	$\eta_i = \mu_i$
Binomial	$\eta_i = \log \left( \frac{\mu_i}{1 - \mu_i} \right)$
Poisson	$\eta_i = \log \mu_i$
Gama	$\eta_i = \mu_i^{-1}$

**Fonte: McCullagh & Nelder (1989).**

### 2.3 Modelos Lineares Generalizados Mistos (MLGM)

Uma extensão natural dos MLG são os modelos lineares generalizados com efeitos fixos e aleatórios denominados MLGM; estes modelos surgem da necessidade de realizar inferências sobre a população. Para tal, considera-se que as variáveis coletadas nos níveis de determinados fatores podem ser considerados pertencentes a uma população de níveis destes fatores (McCulloch & Searle, 2001).

Com a incorporação dos efeitos aleatórios, a distribuição de  $y$  deve ser condicionada a tais efeitos:

$$y_i | u \sim indep. f_{y_i|u}(y_i | u)$$
$$f_{y_i|u}(y_i | u) = \exp \left\{ \frac{1}{\phi} [y_i \theta_i - b_i(\theta_i)] + c_i(y_i, \phi) \right\}, \quad (2.6)$$

em que  $E[y_i | u_i] = \mu_i$  é um termo associado aos efeitos aleatórios e deve ser incorporado no preditor linear como segue:

$$\eta_i = g(\mu_i) = x_i' \beta + z_i' u \Rightarrow \mu_i = g_i^{-1}(\eta_i), \quad (2.7)$$

em que  $Z = (z_1, \dots, z_q)^t$  é uma matriz de incidência para efeitos aleatórios e  $u$  um vetor de parâmetros desconhecidos de dimensões  $(q \times 1)$ .

Como  $u$  é um vetor de efeitos aleatórios deve, ser assumida uma distribuição de probabilidade, qual seja:

$$u \sim f_U(u). \quad (2.8)$$

## 2.4 Inferência Bayesiana

A estimação em MLGM, como em outros modelos pode ser realizada através Máxima Verossimilhança e Quase Máxima Verossimilhança (McCoulch & Searle, 2001), mas um outro procedimento que vem sendo muito difundido é o da teoria bayesiana.

O conhecimento prévio sobre determinado problema não deve ser desprezado em qualquer área do conhecimento. Com relação à análise de dados não é diferente, e informações sobre parâmetros os quais se deseja estimar são introduzidas através das distribuições a priori em que o valor do parâmetro é pensado como uma variável aleatória. Assim, um dos objetivos da inferência bayesiana é acrescentar informações na distribuição amostral de forma a obter melhores estimativas através da distribuição a posteriori.

$$p(\theta | y) \propto p(\theta)p(y | \theta), \quad (2.9)$$

em que  $\theta$  é o parâmetro ou vetor de parâmetros,  $p(\theta | y)$  a distribuição a posteriori,  $p(\theta)$  a priori e  $p(y | \theta)$  a distribuição amostral ou verossimilhança.

A justificativa para a distribuição de probabilidade em (2.9) é feita através do teorema de Bayes:

$$p(\theta | y) = p(y | \theta)p(\theta) / p(y)$$

e

$$p(\theta | y) \propto p(y | \theta)p(\theta), \quad (2.10)$$

já que a função  $p(y)$  da amostra é independente dos parâmetros.

Realizam-se as inferências sobre a distribuição de probabilidade conjunta a posteriori e, na maioria dos casos, deseja-se encontrar uma distribuição para um parâmetro específico. A forma de encontrar tal distribuição, chamada distribuição marginal, é integrando a distribuição posteriori conjunta, obtendo uma função dessa integral para o parâmetro de interesse  $\theta_i$ , isto é,

$$f(\theta_i) = \int \dots \int f(\theta_i, \theta_{-i}) d\theta_{-i}, \quad (2.11)$$

em que  $\theta_{-i}$  é o conjunto complementar de parâmetros para  $\theta_i$ .

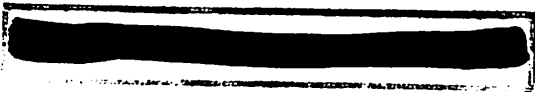
O que ocorre porém, em muitos casos, é que a forma analítica das marginais é complexa e torna impossível ou inviável a integração em (2.11). Este foi, inclusive, o motivo pelo qual a inferência bayesina ficou esquecida um grande período. Nesses casos, a amostragem Gibbs é uma boa alternativa para obtenção de  $f(\theta_i)$ , método inicialmente utilizado por Gelfand et al. (1990), na estatística, o qual foi elaborado por Geman & Geman (1984).

Assim, métodos numéricos são usados, como o método de Monte Carlo via Cadeias de Markov (MCMC), para gerar valores de uma distribuição condicional a posteriori para cada parâmetro.

#### 2.4.1 Priori não informativa

Quando a verossimilhança é dominante ou deseja-se representar o desconhecimento sobre  $\theta$ , a primeira idéia é pensar em todos os possíveis valores para  $\theta$  como igualmente prováveis:

$$p(\theta) \propto k, \quad (2.12)$$



com  $\theta$  limitado em um intervalo real.

Muitas vezes a priori é a forma de se quantificar a incerteza sobre o experimento, porém a priori não-informativa não representa necessariamente o desconhecimento do pesquisador sobre o experimento, mas deve ser usada também de forma a viabilizar a inferência a posteriori ( Box & Tiao, 1992).

Uma outra classe de prioris não informativas geralmente usadas para solução de problemas de escala, foi proposta por Jeffreys (1961), em que se usa a medida de informação esperada de Fisher de  $\theta$  através de Y:

$$I(\theta) = E[(d^2 \log\{p(y | \theta)\} / d\theta^2 | \theta)] . \quad (2.13)$$

A priori não informativa de Jeffreys tem função de densidade da por

$$P(\theta) \propto [I(\theta)]^{1/2}. \quad (2.14)$$

Se  $\theta$  for uma vetor, então

$$p(\theta) \propto \det|I(\theta)|^{1/2}. \quad (2.15)$$

#### 2.4.2 – Procedimentos Bayesianos para Modelos Lineares Generalizados Mistos

Na presença de efeitos aleatórios a função ligadora é reescrita como

$$g(\mu_i) = \eta_i = x_i' \beta + z_i' u , \quad (2.16)$$

em que  $x_i$  e  $z_i$  são as colunas das matrizes de delineamentos de efeitos fixos e aleatórios, respectivamente.

Segundo Sun et al (2000), inferência bayesiana para MLGM pode ser implementada pelo método Monte Carlo via Cadeias de Markov (MCMC) tal como a amostragem Gibbs.

Outros autores, como Gamermam (1996), também utilizam o método MCMC para amostragem da densidade a posteriori e, para os casos em que não existem formas específicas para as distribuições condicionais completas, a amostragem Gibbs pode ser implementada com o auxílio do algoritmo Metropolis & Hastings (Metropolis et al., 1953 e Hastings, 1970), para realizar o sorteio na condicional completa da amostragem Gibbs. Natarajan & Kass (2000) utilizam o Amostrador de Gibbs como um método natural para simulação da posteriori em MLGM.

A motivação para o uso da amostragem Gibbs é que a real distribuição pode ser aproximada por uma distribuição empírica de  $n$  valores de tal forma que  $n$  seja grande o suficiente para que a amostragem Gibbs atinja a convergência (Zeger & Karin, 1991). Gelfand et al. (1990) utilizam a mesma idéia ilustrando a inferência bayesiana em um conjunto de dados normais, usando o Amostrador de Gibbs. Casella & George (1992) utilizam o Amostrador de Gibbs para ilustrar o fato de que as condicionais, quando  $n \rightarrow \infty$  determina as marginais, isto é, quando se tem uma amostra suficientemente grande de um determinado parâmetro dados os demais, temos uma distribuição empírica que se aproxima suficientemente da distribuição marginal.

Yang et al. (1999) comparam o Algoritmo EM e o Amostrador de Gibbs no estudo de modelos lineares generalizados multivariados com mistura finita e verificam que o Amostrador de Gibbs é mais acurado em suas estimativas e apresenta menor desvio padrão, apesar de as diferenças serem pequenas.

### 2.4.3 Amostrador de Gibbs

Como descrito em 2.7, este algoritmo fornece uma forma alternativa para gerar sucessivos valores de uma distribuição condicional completa.

Suponha uma distribuição  $\pi(\theta)$  em que  $\theta = (\theta_1, \dots, \theta_p)$ ; no contexto da inferência bayesiana, o interesse está em amostrar valores das densidades condicionais completas a posteriori  $\pi_i(\theta_i | \theta_{-i}), i = 1, \dots, p$ .

Segundo Gamerman (1996), o algoritmo pode ser descrito através dos seguintes passos:

1 - inicie o contador de iterações da cadeia  $j = 1$  e escolha valores iniciais arbitrários

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)});$$

2 - obtenha um novo valor  $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_p^{(j)})$  a partir de  $\theta^{(j-1)}$  através de sucessivas gerações de valores

$$\begin{aligned} \theta_1^{(j)} &\sim \pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_p^{(j-1)}) \\ \theta_2^{(j)} &\sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)}) \\ &\vdots \\ \theta_p^{(j)} &\sim \pi(\theta_p | \theta_1^{(j)}, \dots, \theta_{p-1}^{(j)}); \end{aligned} \tag{2.17}$$

3 - mude o contador  $j$  para  $j + 1$  e retorne ao passo 2 até a convergência ser alcançada.

Como descrito em 2.4.2, um valor gerado da condicional só pertencerá à distribuição de interesse (marginal) quando o número de iterações tender a infinito. Então, toma-se o valor  $j$  suficientemente grande para que se tenha a convergência.

#### 2.4.4 Amostragem por rejeição com Algoritmo Metropolis & Hastings

Nos casos em que a distribuição condicional a posteriori não tem uma forma fechada, geram-se valores em duas etapas. Na primeira etapa, utiliza-se uma distribuição conhecida como auxílio. A segunda etapa consiste em usar um procedimento para que os valores amostrados sejam realmente representativos da distribuição a posteriori de interesse.

Seja uma densidade auxiliar  $q(\theta)$  da qual é possível gerar valores, e seja também uma constante  $K$  finita tal que  $p(\theta | x) < [Kq(\theta)]$ . O método consiste em gerar um valor  $\theta^*$  da distribuição auxiliar  $q$  e este valor é aceito como sendo da distribuição de interesse, com probabilidade  $\frac{p(\theta^* | x)}{Kq(\theta^*)}$ . Caso contrário,  $\theta^*$  não é aceito como um valor gerado da distribuição e o processo é repetido até que um valor seja aceito.

Este procedimento pode ser representado pelos seguintes passos:

1 – gera um valor  $\theta^*$  da distribuição auxiliar  $q$  ;

2 – gera  $u \sim U(0,1)$  ;

3 – aceita  $\theta^*$  como um valor da densidade de interesse se  $u < \frac{p(\theta^* | x)}{kq(\theta^*)}$ , caso

contrário, rejeita  $\theta^*$  e retorna-se ao passo 1.



## 2.4.5 Diagnósticos de convergência

Existem métodos informais para o diagnóstico de convergência que foram inicialmente propostos por Gelfand et al. (1990). A técnica gráfica que consiste em observar a trajetória da cadeia ao longo das iterações é um dos métodos propostos por eles. Nesta técnica, se o gráfico após um período inicial apresenta repetidamente o mesmo comportamento qualitativo e quantitativo, então pode se concluir pela convergência da cadeia.

Outros métodos existentes são os formais, como por exemplo, o método proposto por Gelman & Rubin (1992), em que, para cada valor  $\psi$  a ser estimado, geram-se  $J$  cadeias paralelas de tamanho  $n$  e calcula-se a variância entre (E) e dentro (D) para estas cadeias.

$$E = \frac{n}{J-1} \sum_{j=1}^J (\bar{\psi}_{.j} - \bar{\psi}_{..})^2 \quad (2.18)$$

sendo  $\bar{\psi}_{.j} = \frac{1}{n} \sum_{i=1}^n \psi_{ij}$ ,  $\bar{\psi}_{..} = \frac{1}{J} \sum_{j=1}^J \bar{\psi}_{.j}$  e

$$D = \frac{1}{J} \sum_{j=1}^J s_j^2, \quad (2.19)$$

sendo  $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\psi_{ij} - \bar{\psi}_{.j})^2$

A necessidade de gerar  $J$  cadeias está justificada pelo uso da variância entre cadeias (E) na verificação da convergência.

Desta forma, pode se estimar a variância marginal a posteriori  $\text{var}(\psi | y)$  pela ponderação de D e E

$$\widehat{\text{var}}^+(\psi | y) = \frac{n-1}{n}D + \frac{1}{n}E \quad (2.20)$$

No limite quando  $n \rightarrow \infty$ , o valor esperado de D aproxima-se de  $\text{var}(\psi | y)$  e o monitoramento da convergência é feita através do fator de redução potencial escalar  $\widehat{R}$ , dado por

$$\widehat{R} = \frac{\widehat{\text{var}}^+(\psi | y)}{D} \quad (2.21)$$

que tende a 1 quando  $n \rightarrow \infty$ . Se o fator  $\widehat{R}$  for superior a 1,2, têm-se razões para afirmar que inferências sobre a distribuição amostrada podem não ser confiáveis.

Existem outros métodos também que podem ser usados, tais como o de Raftery & Lewis (1992), o qual sugere os tamanhos dos saltos que devem ser dados para extrair o efeito do chute inicial (descarte) e para retirar o efeito de dependência entre os valores amostrados (pulo).

## 2.5 MLGM Hierárquico para dados Poisson

Segundo Sun et al. (2000), de acordo com a definição geral de modelos lineares generalizados e fazendo na expressão (2.5)  $\theta = \eta$ , dado  $Y_1..Y_N$ , observações aleatórias independentes,  $Y_i$  tem densidade de probabilidade como segue:

$$f_i(y_i | \eta_i, \phi) = \exp \left\{ \frac{1}{\phi} [y_i \eta_i - b_i(\eta_i)] + c_i(y_i, \phi) \right\}. \quad (2.22)$$

Para o caso em que  $Y_i$  tem distribuição Poisson,  $\phi = 1$ ,  $b_i(\eta_i) = \exp(\eta_i)$  e  $c_i(y_i, \phi) = -\log(y_i!)$ , o modelo acima pode ser reescrito como

$$y | \eta \sim \text{Poisson}(\exp[\eta])$$

$$f(y | \eta) = \exp \{ y\eta - \exp(\eta) - \log(y!) \}, \quad (2.23)$$

com o preditor linear dado como

$$\eta = X\beta + Zu, \quad (2.24)$$

em que  $\eta$  é um vetor de preditores lineares ( $n \times 1$ ) associado ao parâmetro  $\exp(\eta)$ .

É comum o uso de priori normal para os efeitos, mas supondo que não haja informação sobre os efeitos fixos, admite-se uma priori constante caracterizando tal desconhecimento, isto é,  $p(\beta) \sim k$ ; para os componentes de variância  $\sigma_i^2$  assume-se como priori uma gama inversa com parâmetros  $(a_i, b_i)$ ,

$$p_i(\sigma_i^2) \propto \frac{1}{(\sigma_i^2)^{a_i+1}} \exp(-b_i / \sigma_i^2), \quad (2.25)$$

com  $b_i > 0$ . É preciso notar que outros autores têm preferido apresentar prioris para a herdabilidade ao invés das componentes da variância, pois estas prioris

poderiam ser elicitadas mais facilmente pelo pesquisador (Bink et al., 1998), já que é mais fácil assumir que se conhece  $h^2$  que as componentes da variância.

Na Figura 1 estão apresentadas as distribuições a priori possíveis para os coeficientes de herdabilidade para seleção entre cruzamentos

$$h_m^2 = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_p^2/3 + \sigma_e^2/30} \text{ e para seleção individual } h_d^2 = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_p^2 + \sigma_e^2}$$

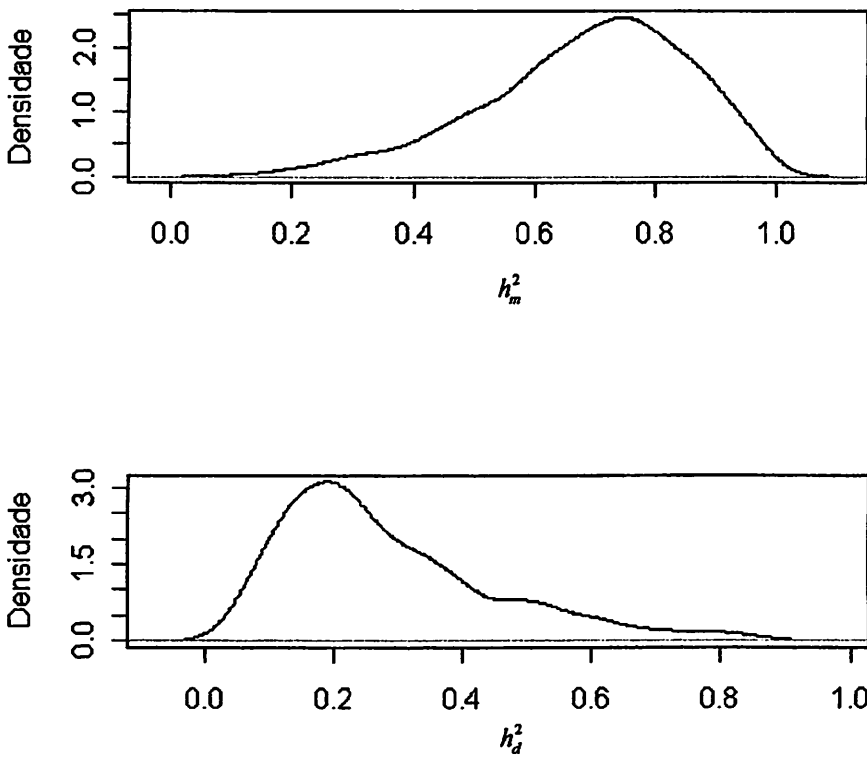


Figura 1: Distribuições de herdabilidades para seleção entre cruzamentos ( $h_m^2$ ) e entre indivíduos ( $h_d^2$ ). (Ver página 63).

Assim, de posse da distribuição amostral e das prioris a distribuição conjunta posteriori dos parâmetros do modelo (2.16) é

$$\begin{aligned} p(\eta, \beta, u, \sigma_u^2, \sigma_e^2 | y) &\propto p(\eta, \beta, u, \sigma_u^2, \sigma_e^2) L(y | \eta, \beta, u, \sigma_u^2, \sigma_e^2 | y) \\ &\propto p(\eta) p(\beta) p(u) p(\sigma_u^2) p(\sigma_e^2) p(y | \eta, \beta, u, \sigma_u^2, \sigma_e^2 | y) \end{aligned} \quad (2.26)$$

As distribuições condicionais completas a posteriori para cada parâmetro do modelo dado em (2.16) são obtidas em (2.26), como segue (Sorensen, 1996):

$$\pi(\theta_i | \theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}, \quad (2.27)$$

em que  $\theta = (\eta, \beta, u, \sigma_u^2, \sigma_e^2)$ ; assim, a posteriori é obtida fazendo

$$\pi(\theta_i | \theta_{-i}) \propto \pi(\theta), \quad (2.28)$$

pois o denominador em (2.27) é constante para o parâmetro de interesse  $\theta_i$ .

Quando a expressão em (2.28) possuir fatores independentes de  $\theta_i$ , será simplificada com relação a tais fatores e a proporcionalidade garante a validade da expressão obtida para a condicional a posteriori completa.

## **3 MATERIAL E MÉTODOS**

### **3.1 Material**

O material vegetal utilizado na exemplificação consistia de um experimento do Programa de Melhoramento Genético de Batata da Universidade Federal de Lavras (UFLA), relatado por Diniz (2002).

Foi utilizado um delineamento experimental em látice 5x5 triplo com 2 testemunhas clonais (tratamentos adicionais) por bloco, num total de 105 parcelas com estande inicial de 30 plantas. Em cada parcela foram coletados, de cada planta, os números de tubérculos graúdos, sendo esta a variável usada para a ilustração.

Os tratamentos avaliados estão dispostos na Tabela 2 e os tratamentos adicionais foram as variedades Bintje e Achat.

A variância residual para as testemunhas é, portanto, associada apenas a fatores ambientais e a variância residual para as parcelas dos demais tratamentos apresenta componentes ambientais e genéticos (variabilidade dos clones dentro de uma mesma família).

Neste sentido, foram realizadas três análises independentes dos experimentos: a) experimento completo, com todos os tratamentos descritos acima; b) análise sem as parcelas referentes às testemunhas e c) análise apenas com as parcelas das testemunhas. O objetivo desta distinção foi obter estimativas para as componentes da variância com diferentes significados (ambiental e genético). Assim, na análise (b) a variância do erro dentro de parcelas era devida às componentes genética e ambiental, enquanto na análise (c) devia-se exclusivamente à componente ambiental.

TABELA 2: Tratamentos avaliados no experimento.

Família	Genealogia
MCR-1	Panda x Aracy
MCR-2	EOA 2 x Premiere
MCR-3	Contenda x Chiquita
MCR-4	XY 12 x XY 13
MCR-5	Chiquita x Atlantic
MCR-6	XY 2 x XY 13
MCR-7	XY 10 x XY 13
MCR-8	Chiquita x Aracy
MCR-9	XY 7 x XY 9
MCR-10	ESL 58 x Aracy
MCR-11	XY 3 x XY 9
MCR-12	XY 9 x XY 13
MCR-13	XY 17 x XY 9
MCR-14	XY 5 x XY 9
MCR-15	XY 9 x XY 19
MCR-16	XY 9 x XY 4
MCR-17	XY 9 x XY 10
MCR-18	Atlantic x Aracy
MCR-19	Panda x Atlantic
MCR-20	Contenda x Aracy
MCR-21	EOA252 x Bulk
MCR-22	XY 2 x XY 4
MCR-23	EOA 256 x Bulk
MCR-24	Panda x Chiquita
MCR-25	XY 2 x XY 3
Achat*	A1 x A2
Bintje*	B1 x B2

Fonte: Diniz, (2002), \* Genealogia não conhecida.



### 3.2 Metodologia

No modelo linear, para fins de análise, não foi considerado isoladamente o efeito de repetições, dado que o fator bloco foi considerado fixo e os efeitos de blocos e repetições foram então confundidos. A variável resposta analisada foi o “número de tubérculos graúdos”. Assumiu-se que esta contagem segue uma distribuição Poisson e que o modelo correto para o experimento em blocos incompletos seria de efeitos multiplicativos. Desta forma, estaria justificado analisar o experimento como um modelo linear generalizado misto, com preditor linear de distribuição normal e resposta Poisson.

#### 3.2.1 Modelo Linear Generalizado Misto Poisson

Usando a função ligadora logarítmica para o MLGM Normal-Poisson, temos o seguinte modelo linear:

$$\eta_{ijk} = \log(\mu_{ijk}) = m + b_i + c_j + p_k, \quad (3.1)$$

em que  $\eta_{ijk}$  é o preditor linear associado ao parâmetro  $\exp(\eta_{ijk})$ ;  $m$  é uma constante associada a todas as observações;  $b_i$  é o efeito do bloco  $i$ ,  $i = 1, 2, \dots, 15$ ;  $c_j$  é o efeito do cruzamento  $j$ ,  $j = 1, 2, \dots, 27$ ; e  $p_k$  é o efeito da parcela  $k$ ,  $k = 1, 2, \dots, 105$ . Na forma matricial, pode-se escrever o modelo em 3.1 como segue:

$$\begin{aligned} E[y] &= \mu \\ \eta &= \log(\mu) = X\beta + Zu, \end{aligned} \quad (3.2)$$



em que  $y$  é um vetor ( $n \times 1$ ) de observações, neste caso número de tubérculos graúdos;  $\eta$  é o vetor dos preditores lineares, de dimensão  $n \times 1$ ;  $X$  é a matriz de delineamento para efeitos fixos, de dimensão  $n \times p$ ;  $Z$  é a matriz do delineamento para efeitos aleatórios, de dimensão  $n \times q$ ;  $\beta$  é o vetor de parâmetros para efeitos fixos  $p \times 1$ , no presente caso efeito de blocos, e  $u$  é o vetor de parâmetros para efeitos aleatórios  $q \times 1$ .

O modelo linear generalizado misto foi ajustado para os dados descritos em 3.1, com um programa implementado na linguagem R 1.7.1, conforme Anexo C, em que o número de tubérculos graúdos foi analisado conforme o modelo dado por (3.2). Nesse caso específico, considerou-se o vetor  $u = [u_p : u_c]$ , ou seja, particionado em efeitos de parcelas e genótipos.

Com base no modelo 3.2 e sendo as amostras realizadas nas parcelas, a variância dentro, correspondente à variância residual ( $\sigma_e^2$ ), pode possuir efeito ambiental e/ou genético conforme os tratamentos considerados.

Foram, então, calculadas estimativas de herdabilidade para seleção entre cruzamentos e dentro de parcelas, utilizando as seguintes expressões:

i) herdabilidade entre cruzamentos

$$\hat{h}_m^2 = \frac{\hat{\sigma}_c^2}{\hat{\sigma}_c^2 + \frac{\hat{\sigma}_p^2}{r} + \frac{\hat{\sigma}_e^2}{kr}}, \quad (3.3)$$

em que  $\hat{\sigma}_c^2$  e  $\hat{\sigma}_p^2$  são as estimativas das componentes da variância entre cruzamentos e entre parcelas, respectivamente,  $r$  é o número de repetições do ensaio e  $k$  é o número de parcelas.

ii) herdabilidade para seleção dentro de parcelas

$$\hat{h}_{nd}^2 = \frac{\hat{\sigma}_{gd}^2}{\hat{\sigma}_{gd}^2 + \hat{\sigma}_{ed}^2} = \frac{\hat{\sigma}_e^{2*} - \hat{\sigma}_e^{2**}}{\hat{\sigma}_e^{2*}}, \quad (3.4)$$

em que  $\hat{\sigma}_e^{2*}$  é a variância residual estimada no experimento apenas com cruzamentos e  $\hat{\sigma}_e^{2**}$  é a variância residual no experimento apenas com as testemunhas.

### 3.2.2 Inferência Bayesiana para Modelo Linear Generalizado Misto Poisson

Para a obtenção da posteriori conjunta, considerou-se a verossimilhança da família exponencial, para o caso específico Poisson, será usada a seguinte notação:

$y \cdot \eta$  é o produto de Hadamard entre os elementos dos vetores  $y$  e  $\eta$ ;

$$\exp(\eta) = \begin{bmatrix} e(\eta_1) \\ e(\eta_2) \\ \vdots \\ e(\eta_n) \end{bmatrix}, \text{ em que } \exp(\eta) \text{ é a exponencial de cada elemento do vetor } \eta,$$

$$\text{e } \log(y!) = \begin{bmatrix} \log(y_1!) \\ \log(y_2!) \\ \vdots \\ \log(y_n!) \end{bmatrix}, \text{ em que } \log(y!) \text{ é o logaritmo na base } e \text{ de cada}$$

elemento do vetor  $y$ . Assim, fazendo a forma geral, tem-se:

$$f(y | \eta) = \exp\{y \cdot \eta - \exp(\eta) - \log(y!)\} \quad (3.5)$$

Para cada parâmetro foram estabelecidas prioris não informativas, tais como:

i) inversa gama para as componentes da variância

$$g_i(\sigma_i^2) = \frac{1}{\sigma_i^{2(a_i+1)}} \exp(-b_i / \sigma_i^2), \quad i = e, p \text{ e } c, \quad (3.6)$$

em que  $\sigma_i^2$  é a componente da variância associada ao efeito aleatório "i" (resíduos, parcelas e genótipos) e  $a_e = 10$ ,  $a_p = 1$ ,  $a_c = 1$ ,  $b_e = 5$ ,  $b_p = 5$  e  $b_c = 5$  são os parâmetros das prioris para cada componente da variância. Estas prioris são na verdade bastante vagas, tendo influência mínima nos resultados.

A Figura 2 apresenta os gráficos das distribuições representadas por estas prioris. Isto representa a incerteza sobre as componentes da variância antes de realizar a análise.

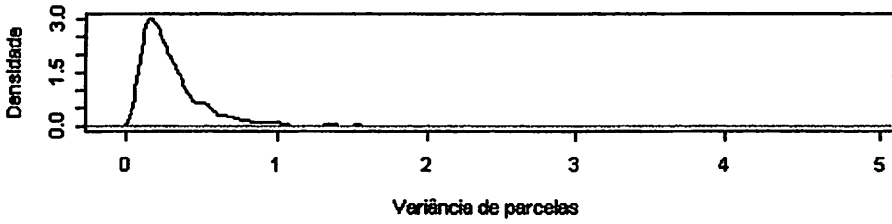
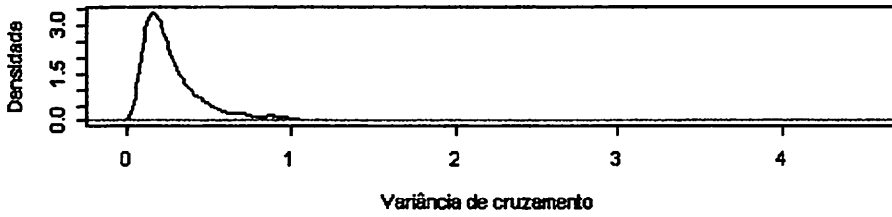
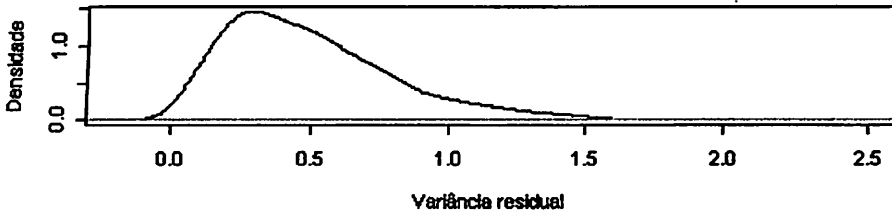


Figura 2: Distribuições a priori para as componentes da variância residual ( $\sigma_e^2$ ), de cruzamentos ( $\sigma_c^2$ ) e de parcelas ( $\sigma_p^2$ ).

ii) Constantes (uniformes) para os efeitos fixos

$$p(\beta) \propto k \tag{3.7}$$

A escolha desta priori para os efeitos fixos é conveniente pois representa a ausência de conhecimento sobre as médias, situação análoga à da análise tradicional dos melhoristas.

iii) normais (vagas) para efeitos aleatórios

$$p(u) \propto \left(\frac{1}{\sigma_p^2}\right)^{\frac{q_p}{2}} \left(\frac{1}{\sigma_c^2}\right)^{\frac{q_c}{2}} \exp\left\{-\frac{1}{2}(u'U^{-1}u)\right\}, \quad (3.8)$$

sendo

$$U = \begin{pmatrix} P & \emptyset \\ \emptyset & G \end{pmatrix} \quad (3.9)$$

$$P = \sigma_p^2 I_{(105)}, \quad G = \sigma_c^2 I_{(27)} \quad \text{e } u = [u_p : u_c],$$

em que  $\sigma_p^2$ ,  $\sigma_c^2$ ,  $u_p$  e  $u_c$  são os componentes da variância e vetores de efeitos para parcelas e genótipos, respectivamente.

Desta forma, a distribuição a priori para efeito aleatório foi assumida como:

$$u|U \sim N(\emptyset, U)$$

Isto indica uma escolha bastante vaga para a distribuição a priori, o que garante que a análise tenha pouca influência de qualquer conhecimento a priori. Isto tudo serve como garantia de uma boa comparação entre os resultados da análise tradicional e da bayesiana, mas uma potencial vantagem da análise bayesiana seria justamente incorporar o eventual conhecimento sobre as médias numa distribuição a priori mais informativa.

### 3.2.3 Posteriores conjunta

A distribuição a posteriori conjunta, de acordo com as distribuições a priori em (3.6), (3.7) e (3.8), e a verossimilhança em (3.5) podem ser escritas de forma proporcional:

$$p(\eta, \beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2 | y) \propto p(\beta) p(\sigma_p^2) p(\sigma_c^2) p(\sigma_e^2) p(u) L(y | \eta, \beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2), \quad (3.10)$$

ou seja,

$$\begin{aligned} p(\eta, \beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2 | y) &\propto \left(\frac{1}{\sigma_e^2}\right)^{q_e+1} \exp\left\{-\frac{b_e}{\sigma_e^2}\right\} \left(\frac{1}{\sigma_p^2}\right)^{q_p+1} \exp\left\{-\frac{b_p}{\sigma_p^2}\right\} \\ &\times \left(\frac{1}{\sigma_c^2}\right)^{q_c+1} \exp\left\{-\frac{b_c}{\sigma_c^2}\right\} \left(\frac{1}{\sigma_u^2}\right)^{\frac{q}{2}} \exp\left\{-\frac{1}{2}(u'U^{-1}u)\right\} \\ &\times \left(\frac{1}{\sigma_e^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2}(\eta - X\beta - Zu)'(\eta - X\beta - Zu)\right\} \\ &\times \exp\{y \circ \eta - \exp(\eta) - \log(y!)\} \end{aligned} \quad (3.11)$$

em que  $q = q_p + q_c$  denota a quantidade de genótipos e de parcelas e  $n$ , a quantidade de observações. Os termos da exponencial e logaritmo envolvidos estão definidos nas páginas 24 e 25.

A partir desta posteriori conjunta foram obtidas as distribuições condicionais a posteriori completas para cada parâmetro, nas quais foi aplicada a amostragem Gibbs com o intuito de obter aproximações das marginais, como dado a seguir.

### 3.2.4 Forma de obtenção das condicionais completas

Segundo Sorensen (1996), dado um vetor de parâmetros  $\theta = (\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_n)$  com distribuição estacionária  $\pi(\theta)$ , cada elemento que pode ser um vetor ou um escalar terá distribuição condicional completa dada por

$$\pi(\theta_i | \theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}, \quad (3.12)$$

em que  $\theta_{-i}$  é o conjunto complementar de parâmetros para  $\theta_i$ .

No presente caso, considerou-se  $\theta = (\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta)$ .

Assim, para o caso da distribuição condicional completa para  $\beta$ , tem-se:

$$p(\beta | u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta, y) = \frac{p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y)}{\int p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y) d\beta}. \quad (3.13)$$

Como a integral no denominador em (3.12) não depende de  $\beta$  e os termos independentes de  $\beta$  no numerador são constantes para este parâmetro, a expressão pode então ser escrita de forma proporcional.

Os resultados referentes a cada uma das condicionais completas são apresentados na seção 4.2 do desenvolvimento.

## 4 DESENVOLVIMENTO

### 4.1 Análise exploratória

Inicialmente foram apresentadas algumas estatísticas descritivas para as distribuições do número de tubérculos graúdos em relação aos diferentes fatores experimentais (em especial, para os cruzamentos), por meio de Box Plots. Logo após, a metodologia descrita na seção 3 foi aplicada aos dados, conforme a estrutura do delineamento apresentado. Os resultados e alguns comentários que podem ilustrar pontos de interesse para a aplicação da inferência bayesiana no exemplo utilizado estão apresentados a seguir.

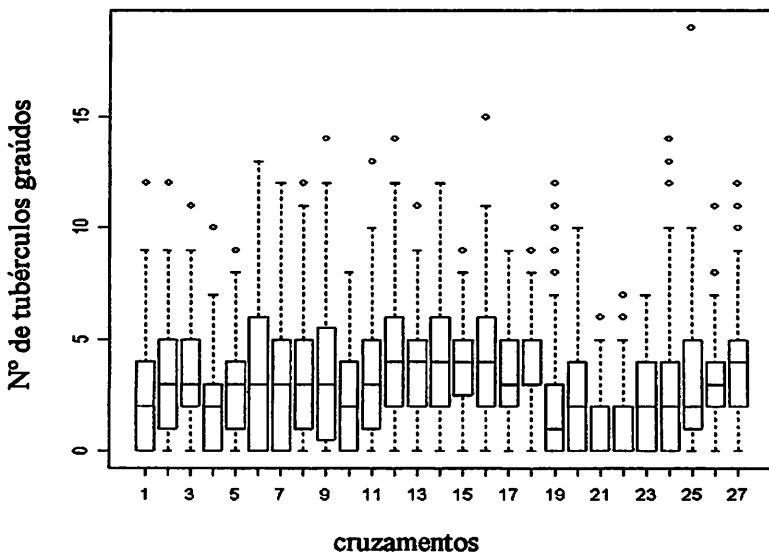


Figura 3: Box Plot para a característica número de tubérculo graúdos em cada cruzamento.



Observando a Figura 3, pode-se verificar que as distribuições para os cruzamentos apresentam forte assimetria, justificada pela presença de outliers no sentido de aumentar a expressão da variável praticamente em todos os casos.

É importante notar que as diferenças existentes na variabilidade dentro de cada cruzamento não são discrepantes ao ponto de serem detectadas em testes para verificação de heterogeneidade das variâncias.

Portanto, pode-se concluir que as médias dos cruzamentos seguem uma distribuição de probabilidade e assim, justifica-se considerar cruzamentos como efeitos aleatórios.

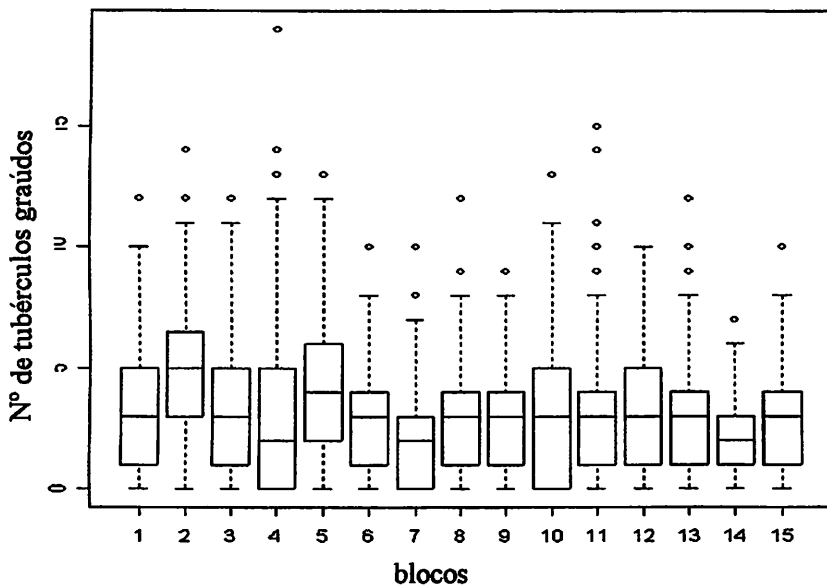


Figura 4: Box Plot para a característica número de tubérculo graúdos em cada bloco.

Para o fator bloco, verifica-se também número de tubérculos graúdos discrepantes em quase todos os casos, distribuição assimétrica e variabilidades semelhantes dentro de cada um, conforme Figura 4.

Neste caso, testes para verificação de heterogeneidade das variâncias também não detectarão diferenças significativas entre blocos.

Verifica-se que bloco também tem justificativa para ser considerado como efeito aleatório; porém, por questão metodológica, foi considerado como efeito fixo.

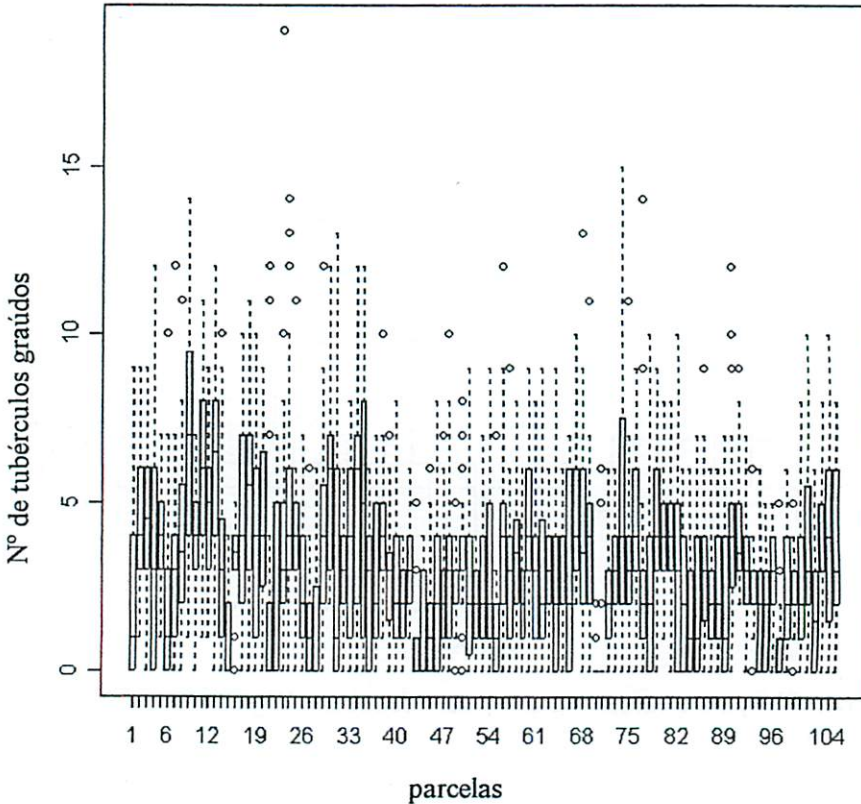


Figura 5: Box Plot para a característica número de tubérculo graúdo em cada parcela.

Na Figura 5 observa-se, como para os demais fatores, que a variável analisada em cada parcela apresenta valores que se destacam quanto à contagem, ou seja, apresenta forte assimetria, justificada pela presença de outliers no sentido de aumentar a expressão da variável, praticamente em todas as parcelas.

A variância de parcelas é considerada homogênea; e portanto, conclui-se que as médias das parcelas seguem uma distribuição de probabilidade e, assim, justifica-se considerá-las como de efeitos aleatórios.

#### 4.2 Obtenção das condicionais completas

A condicional completa para efeitos fixos de blocos segundo as equações (3.13) e (3.11) das páginas 29 e 28 é dada como:

$$\begin{aligned}
 p(\beta | u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta, y) &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} (\eta - X\beta - Zu)' (\eta - \beta X - Zu) \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} (\eta - X\beta - Zu)' (\eta - X\beta - Zu) \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} [(\eta - Zu)' - (X\beta)'] [(\eta - Zu) - (X\beta)] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} [(\eta - Zu)' (\eta - Zu) - ((\eta - Zu)' X\beta - \beta' X' (\eta - Zu) + \beta' X' (X\beta))] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} [ -(\eta - Zu)' X\beta - \beta' X' (\eta - Zu) + \beta' X' (X\beta) ] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} [\beta - (X'X)^{-1} X' (\eta - Zu)]' (X'X) [\beta - (X'X)^{-1} X' (\eta - Zu)] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} [ -(\eta - Zu)' X (X'X)^{-1} X' (\eta - Zu) ] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} [\beta - (X'X)^{-1} X' (\eta - Zu)]' (X'X) [\beta - (X'X)^{-1} X' (\eta - Zu)] \right\}.
 \end{aligned}
 \tag{4.1}$$

Pode ser verificado em (4.1) que a distribuição condicional completa a posteriori para  $\beta$  é uma normal:

$$\beta | (\eta, u, \sigma_e^2, \sigma_p^2, \sigma_c^2, y) \sim N((X'X)^{-1}X'(\eta - Zu), \sigma_e^2(X'X)^{-1}), \quad (4.2)$$

em que  $\beta$  foi reparametrizado para representar as médias de blocos e não os efeitos de blocos; desta forma,  $X'X$  tem inversa clássica.

A posteriori condicional completa de  $u$  pode ser encontrada de forma análoga à distribuição a posteriori condicional completa de  $\beta$ , ou seja, dada a posteriori conjunta (3.11), na página 28, tem-se:

$$\begin{aligned} p(u | \beta, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta, y) &= \frac{p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y)}{\int p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y) du} \\ &\propto \exp \left\{ \frac{1}{2\sigma_e^2} (u' \sigma_e^2 U^{-1} u + (\eta - X\beta - Zu)' (\eta - X\beta - Zu)) \right\} \\ &\propto \exp \left\{ \frac{1}{2\sigma_e^2} (u' \sigma_e^2 U^{-1} u + [(\eta - X\beta)' - (Zu)'] [(\eta - X\beta) - (Zu)]) \right\} \\ &\propto \exp \left\{ \frac{1}{2\sigma_e^2} (u' \sigma_e^2 U^{-1} u - (\eta - X\beta)' Zu - u' Z' (\eta - X\beta) + u' Z' (Zu)) \right\} \\ &\propto \exp \left\{ \frac{1}{2\sigma_e^2} [u - (Z'Z + \sigma_e^2 U^{-1})^{-1} Z' (\eta - X\beta)]' (Z'Z + \sigma_e^2 U^{-1}) [u - (Z'Z + \sigma_e^2 U^{-1})^{-1} Z' (\eta - X\beta)] \right\} \\ &\times \exp \left\{ \frac{1}{2\sigma_e^2} (-[Z' ((\eta - X\beta)'] (Z'Z + \sigma_e^2 U^{-1})^{-1} [Z' (\eta - X\beta)]) \right\} \\ &\propto \exp \left\{ \frac{1}{2\sigma_e^2} [u - (Z'Z + \sigma_e^2 U^{-1})^{-1} Z' (\eta - X\beta)]' (Z'Z + \sigma_e^2 U^{-1}) [u - (Z'Z + \sigma_e^2 U^{-1})^{-1} Z' (\eta - X\beta)] \right\} \end{aligned} \quad (4.3)$$

Assim verifica-se, através de (4.3) que a distribuição condicional completa a posteriori para  $u$  é uma normal:

$$u | (\eta, \beta, \sigma_e^2, \sigma_p^2, \sigma_c^2, y) \sim N(MZ'(\eta - X\beta), \sigma_e^2 M), \quad (4.4)$$

em que  $M = (Z'Z + \sigma_e^2 U^{-1})^{-1}$  e  $U$  é como na página 27.

A distribuição condicional completa para o componente da variância  $\sigma_p^2$ , obtida através de (3.11), é dada a seguir:

$$p(\sigma_p^2 | \beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta, y) = \frac{p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y)}{\int p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y) d\sigma_p^2} \quad (4.5)$$

$$\propto \left( \frac{1}{\sigma_p^2} \right)^{a_p + \frac{q_p}{2} + 1} \exp \left\{ -\frac{1}{\sigma_p^2} \left[ \frac{1}{2} (u_p' J u_p) + b_p \right] \right\}$$

em que a integral no denominador não depende de  $\sigma_p^2$ ; além disso, no numerador, dado por (3.11), existem fatores que são constantes com relação a  $\sigma_p^2$ . Assim, obtém-se proporcionalmente a distribuição a posteriori completa para a componente da variância de parcelas, dada por uma Qui-Quadrado Inversa Escalada:

$$\sigma_p^2 | (\eta, \beta, u, \sigma_e^2, \sigma_c^2, y) \sim Inv - \chi^2 \left( a_p + \frac{q_p}{2}, b_p + \frac{1}{2} u_p' J u_p \right), \quad (4.6)$$

da qual pode-se obter amostras invertendo a variável de uma Qui-Quadrado multiplicada por  $\frac{1}{2}(u_p' I u_p) + b_p$  com parâmetro  $a_p + \frac{q_p}{2}$ .

A distribuição condicional completa para o componente da variância  $\sigma_c^2$ , da mesma forma como foi obtida  $\sigma_p^2$  em (3.11), é dada como segue:

$$p(\sigma_c^2 | \beta, u, \sigma_p^2, \sigma_e^2, \eta, y) = \frac{p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y)}{\int p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y) d\sigma_c^2} \quad (4.7)$$

$$\propto \left( \frac{1}{\sigma_c^2} \right)^{a_c + \frac{q_c}{2} + 1} \exp \left\{ -\frac{1}{\sigma_c^2} \left[ \frac{1}{2} (u_c' I u_c) + b_c \right] \right\}$$

que é também uma Qui-Quadrado Inversa Escalada:

$$\sigma_c^2 | (\eta, \beta, u, \sigma_e^2, \sigma_p^2, y) \sim Inv - \chi^2 \left( a_c + \frac{q_c}{2}, b_c + \frac{1}{2} u_c' I u_c \right) \quad (4.8)$$

com parâmetro  $a_c + \frac{q_c}{2}$  e pode ser amostrada da mesma forma como amostrada a distribuição em (4.6).

Para a componente da variância  $\sigma_e^2$ , a condicional completa a posteriori obtida por meio da expressão (3.11), na página 28, é dada como segue:

$$p(\sigma_e^2 | \beta, u, \sigma_p^2, \sigma_c^2, \eta, y) = \frac{p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y)}{\int p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y) d\sigma_e^2} \quad (4.9)$$

$$\propto \left( \frac{1}{\sigma_e^2} \right)^{a_e + \frac{n}{2} + 1} \times \exp \left\{ -\frac{1}{2\sigma_e^2} [(\eta - X\beta - Zu)' (\eta - X\beta - Zu) + b_e] \right\}$$

isto é,

$$\sigma_e^2 | (\eta, \beta, u, \sigma_p^2, \sigma_c^2, y) \sim \text{Inv-}\chi^2 \left( a_e + \frac{n}{2}, \frac{(\eta - X\beta - Zu)^t (\eta - X\beta - Zu) + b_e}{2} \right), \quad (4.10)$$

a qual pode ser amostrada de forma análoga à amostragem para  $\sigma_p^2$ .

A densidade condicional completa a posteriori para  $\eta$ , obtida também por meio da expressão (3.11), é dada como:

$$p(\eta | \beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, y) = \frac{p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y)}{\int p(\beta, u, \sigma_p^2, \sigma_c^2, \sigma_e^2, \eta | y) d\eta} \\ \propto \exp \{ y\eta - \exp(\eta) - \log(y!) \} \times \exp \left\{ - \frac{(\eta - X\beta - Zu)^t (\eta - X\beta - Zu)}{2\sigma_e^2} \right\}, \quad (4.11)$$

em que a proporcionalidade foi obtida sendo o denominador em (4.11) uma constante para  $\eta$  e no numerador, dado pela conjunta em (3.11); considerou-se apenas os fatores dependentes deste parâmetro.

A densidade em (4.11) não tem uma função da qual se possam tomar diretamente amostras Monte Carlo ordinárias. A solução apresentada na literatura é utilizar, nestes casos, a amostragem por rejeição. Especificamente nesse trabalho foi implementado este passo (Algoritmo Metropolis & Hastings), notando que a densidade é um produto de uma Normal e uma Poisson, como segue:

$$g(\eta) \times f(\eta),$$

em que

$$g(\eta) = \exp[y\eta - \exp(\eta) - \log(y!)] \quad (\text{Poisson}) \quad (4.12)$$

e

$$f(\eta) = \exp\left[-\frac{(\eta - X\beta - Zu)^2 (\eta - X\beta - Zu)}{2\sigma_e^2}\right] \quad (\text{Normal}). \quad (4.13)$$

Assim, a amostragem de  $\eta$  pode ser realizada através do método descrito em 2.7.2, sendo  $f(\eta)$  a distribuição auxiliar, ou seja, da qual um valor  $\eta^*$  é amostrado e aceito com probabilidade  $\frac{g(\eta^*) \times f(\eta^*)}{f(\eta^*)} = g(\eta^*)$ . Verifica-se que  $f$  e  $g$  são distribuições de densidades e, assim  $0 \leq f \leq 1$  e  $0 \leq g \leq 1$ ; então  $f.g \leq f$ , como pode ser verificado nas Figuras 6, 7 e 8, satisfazendo a condição para a amostragem com o passo Metropolis & Hastings, descrito em 2.4.4.

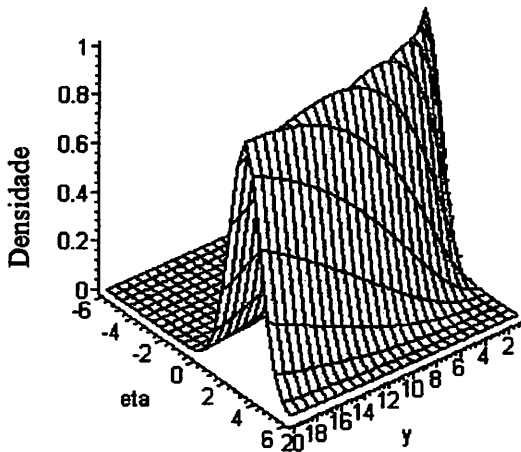


Figura 6: Gráfico da distribuição  $f(\eta)$ .



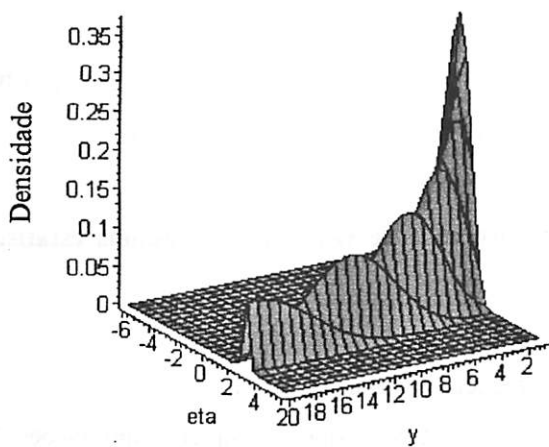


Figura 7: Gráfico da distribuição  $g(\eta)$ .

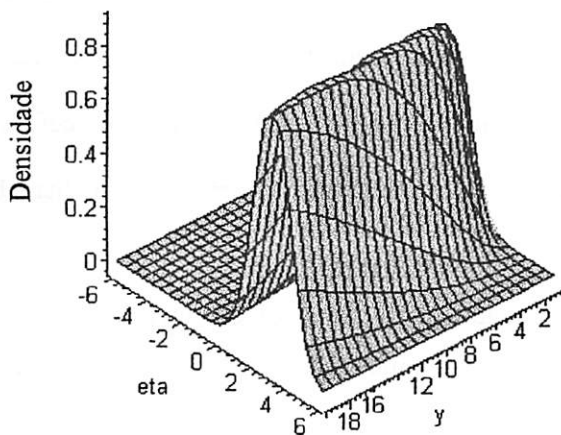


Figura 8: Gráfico da diferença  $(f(\eta)-g(\eta))$ .

A amostragem por rejeição pode ser implementada considerando  $f(\eta)$  como a distribuição auxiliar dado que esta é um “envelope” para a densidade posteriori completa de  $\eta$ , pois  $f(\eta) - g(\eta) > 0$ , conforme Figura 3, ou seja,  $f(\eta) > g(\eta)$ .

O processo amostral foi concluído utilizando a amostragem por rejeição para  $\eta$  através das funções  $f(\eta)$  e  $g(\eta)$  ilustradas na Figura 3.

#### **4.3 Características do Amostrador de Gibbs para algumas estatísticas de interesse**

A vantagem da inferência bayesiana via amostras das distribuições “a posteriori” para o cálculo de estimativas é que quaisquer combinações, lineares ou não lineares, dos valores amostrais, uma vez calculadas em cada passo de amostragem, geram diretamente um ponto amostral para a estatística de interesse, com base na qual a inferência é simples.

Assim, foram geradas cadeias para a variância genética dentro de cada ambiente amostral ( $\sigma_{gd}^2$ ), dada por uma combinação linear entre as variâncias residual do experimento apenas com cruzamentos ( $\sigma_e^{2*}$ ) e a variância residual do experimento apenas com testemunhas ( $\sigma_e^{2**}$ ), e também para o coeficientes de herdabilidade entre cruzamentos ( $h_m^2$ ) e para o coeficiente de herdabilidade dentro de parcelas ( $h_d^2$ ).

Distribuições a posteriori para contrastes de interesse foram obtidas através de combinações lineares dos parâmetros em cada ponto amostral de suas cadeias.

A implementação da amostragem Gibbs no R 1.7.1 está no Anexo C.

#### 4.4 Implementação do processo amostral e diagnósticos de convergência

Foi gerada uma cadeia inicial (com 4.000 iterações) para todos os parâmetros objetivando indicativos de tamanhos de *queima* e *pulo*, calculados pelo critério Raftery & Lewis o qual indicou, dentre todos parâmetros, os maiores valores de 63 e 12 para *queima* (“*burn in*”) e *pulo* (“*thinning*”), respectivamente.

Foram obtidas duas amostras válidas com  $n = 3.800$  observações, a partir de duas cadeias de cem mil (100.000) iterações cada uma, considerando-se a *queima* das 5.000 primeiras iterações e a tomada de pontos amostrais em uma a cada 25 iterações. Isto foi feito para garantir a segurança do processo amostral, apesar de tais valores serem considerados excessivos pelo critério de Raftery & Lewis (1992).

A convergência das cadeias resultantes foi verificada, para cada parâmetro, através dos métodos de Brooks, Gelman & Rubin, implementados na suite de rotinas Bayesian Output Analysis (BOA) do R (Smith, 2003), conforme descrito, verificando-se um fator de redução potencial próximo de 1 ( $\hat{R} \cong 1,133629$ ), valor considerado aceitável.

Na Tabela do anexo A estão dispostos os valores para verificação da convergência da distribuição empírica para cada parâmetro (efeitos fixos, aleatórios e componentes da variância), verificada através do método de Brooks, Gelman & Rubin, que também apresentou o fator de redução potencial próximo de 1 quando  $n \rightarrow \infty$ , e de Raftery & Lewis, o qual, quando aplicado na amostra válida, apresentou, como esperado, *pulo* mínimo para retirada do efeito de dependência entre os valores amostrados e mínimo também quanto à *queima*, indicando ser o processo amostral estável e independente.

Os critérios apresentaram-se coerentes. Na amostra válida já foram *queimados* 5.000 valores iniciais e realizados *pulos* de 25 observações, sendo

esperado, portanto, que os valores indicados pelos critérios de Raftery & Lewis não devessem ser expressivos.

#### 4.5 Estimativas de Componentes da variância

Na Tabela 3 estão apresentadas estimativas da variância de cruzamentos  $\sigma_c^2$ , variância de parcelas  $\sigma_p^2$  e variância residual  $\sigma_e^2$ , calculadas tomando-se médias marginais da distribuição “a posteriori” para três situações de análise, as quais estão descritas em 3.1.

A menor estimativa da  $\sigma_e^2$  foi obtida no experimento considerando apenas as testemunhas. Este resultado é coerente porque o erro dentro de parcelas, neste caso, é devido somente a fatores ambientais, pois neste experimento os genótipos são clones comerciais. Da mesma forma, esperávamos que a estimativa da  $\sigma_e^2$  no experimento completo fosse menor que no experimento só com cruzamentos, devido à melhor precisão nas estimativas, atribuída ao tamanho da amostra.

A Figura 1B ilustra as afirmações acima, com as distribuições completas em gráficos de densidade nuclear normal. A tendência central da densidade para o componente da variância  $\sigma_e^2$  considerando apenas testemunhas é menor que a do experimento completo e esta, por sua vez, é menor do que a do experimento sem testemunhas.

A figura mostra também a dispersão das estimativas em torno da tendência central. A amplitude dos gráficos está associada, neste caso, não só à presença da variabilidade em si, mas essencialmente a diferenças nos tamanhos dos experimentos (e, portanto, diferentes precisões nas estimativas),

A diferença entre as variâncias residuais nos experimentos só com cruzamentos e só com testemunhas são indícios da presença de importante variabilidade genética dentro de parcelas.

As estimativas da variância de cruzamento ( $\sigma_c^2$ ) e da variância de parcela ( $\sigma_p^2$ ) foram menores, em média, no experimento completo, em que o número de cruzamentos é maior (27), seguidas do experimento sem testemunha com 25 cruzamento e do experimento considerando somente testemunhas (Tabela 3)

A menor ou maior estimativa da variância de parcelas está associada, respectivamente, à maior ou menor precisão, pois neste caso o significado genético da variância é parecido, sendo, no entanto, os experimentos de tamanhos bem diferentes. No experimento completo foram consideradas 105 parcelas, portanto graus de liberdade maior, para o erro experimental, comparado com o experimento sem testemunha (75 Parcelas) e com o experimento considerando apenas testemunhas (30 parcelas).

As razões para a maior ou menor estimativa de variância de cruzamentos, no entanto, devem ser encontradas nas próprias estimativas. Em geral, as testemunhas são materiais geneticamente bem distintos que podem ser facilmente interpretáveis como de efeito fixo (variância infinita ou distribuições independentes).

O menor valor para a média da estimativa da variância de cruzamento ( $\sigma_c^2 = 0,3460$ ) ocorreu no experimento completo. A  $\sigma_c^2 = 2,1240$ , estimada no experimento apenas com as testemunhas, é degenerada como se verificou através da Figura 2B, na qual se ilustraram também as densidades para as variâncias de cruzamentos nos demais experimentos.

Os cruzamentos, no entanto apresentam estimativas de valor genético que podem ser agrupadas em uma clara distribuição de probabilidade (mais

agrupada que a normal), conforme pode ser observado na Figura 13B (gráficos de quantis normais).

TABELA 3: Distribuições a posteriori para os componentes de variância associadas ao modelo linear generalizado misto.

	Var	LI	Média	LS	EP	Erro MC
Exp, Completo	$\sigma_c^2$	0,1170	0,3460	0,6389	0,1511	0,0024
	$\sigma_p^2$	0,1136	0,1953	0,2841	0,0451	0,0007
	$\sigma_e^2$	0,2991	0,3387	0,3811	0,0213	0,0006
Sem testemunha	$\sigma_c^2$	0,1510	0,4063	0,7695	0,1869	0,0030
	$\sigma_p^2$	0,1452	0,2818	0,4519	0,0833	0,0014
	$\sigma_e^2$	0,4396	0,5142	0,5861	0,0377	0,0009
Apenas Testemunha	$\sigma_c^2$	0,1340	2,1240	5,5301	15,1075	0,2469
	$\sigma_p^2$	0,1362	0,3621	0,6945	0,1632	0,0032
	$\sigma_e^2$	0,2409	0,2908	0,3415	0,0265	0,0005

#### 4.6 Variância genética e herdabilidade

As estimativas de variâncias de cruzamentos permitem o cálculo da herdabilidade e do ganho pela seleção entre cruzamentos

$$\left( \hat{h}_m^2 = \frac{\hat{\sigma}_c^2}{\hat{\sigma}_c^2 + \frac{\hat{\sigma}_p^2}{r} + \frac{\hat{\sigma}_e^2}{kr}} \right),$$

em que  $r = 3$  (número de repetições do ensaio) e  $k = 30$  o número de plantas por parcela. Adicionalmente, pode-se calcular também, combinando os dados das diferentes formas de análise, estimativas da variância genética dentro de cruzamentos e, desta forma, prever o ganho pela seleção dos melhores genótipos dentro das parcelas dos melhores cruzamentos, usando a seguinte expressão para a herdabilidade:

$$\hat{h}_d^2 = \frac{\hat{\sigma}_{gd}^2}{\hat{\sigma}_{gd}^2 + \hat{\sigma}_{ed}^2} = \frac{\hat{\sigma}_e^{2*} - \hat{\sigma}_e^{2**}}{\hat{\sigma}_e^{2*}},$$

em que  $\hat{\sigma}_e^{2*}$  é a variância residual estimada no experimento apenas com cruzamentos e  $\hat{\sigma}_e^{2**}$  é a variância residual no experimento apenas com as testemunhas. As estimativas de componentes da variância e herdabilidades podem ser encontradas na Tabela 4.

As distribuições a posteriori para estimativas da variância genética dentro de parcelas e herdabilidades  $\hat{h}_m^2$  e  $\hat{h}_d^2$  estão ilustradas na Figura 4B. A variância genética dentro de parcelas apresenta uma distribuição simétrica e as herdabilidades distribuições levemente assimétricas à esquerda. Os resumos destas três densidades estão apresentados na Tabela 4.

A estimativa do coeficiente de herdabilidade entre médias de cruzamentos ( $\hat{h}_m^2$ ) indicou um valor de 82,54% para a característica número de tubérculos graúdos e uma vantagem considerável nesta apresentação é o intervalo de probabilidade obtido com base na distribuição a posteriori para herdabilidade, dada por uma combinação linear de estatísticas adequadas. Para a característica porcentagem de tubérculos graúdos, este valor foi de 70,98% (Diniz, 2002).

TABELA 4: Resumos de estatísticas para variância genética dentro e herdabilidade para seleção entre e dentro de cruzamentos.

	LI	Média	LS	EP	Erro Mc
$\hat{\sigma}_{gd}^2$	0,1329	0,2233	0,3143	0,0462	0,0010
$\hat{h}_m^2$	0,6888	0,8172	0,9286	0,0641	0,0011
$\hat{h}_d^2$	0,3027	0,4313	0,5642	0,0669	0,0014

A herdabilidade na seleção dos melhores genótipos dentro de cruzamentos ( $\hat{h}_d^2$ ), com valor de 43,13%, também apresenta a vantagem descrita para ( $\hat{h}_m^2$ ).

Amaro (2002), em seu trabalho de avaliação da viabilidade da seleção precoce de tubérculos para os caracteres produção (g/planta) e porcentagem de gráudo, em uma segunda geração, encontrou herdabilidades de 86,4% e 78,7%, respectivamente. O autor cita ainda que Souza (1999) determinou herdabilidade de 0,46 entre clones para a produção(g/planta), enquanto Menezes (1999) encontrou, para a mesma variável, herdabilidades variando entre 0,62 a 0,75 entre clones e de 0,73 a 0,97 entre famílias; Lambert (2001) encontrou herdabilidade entre clones variando entre 0,62 a 0,80 e entre famílias variando de 0,83 a 0,96.

As estimativas encontradas na Tabela 4 são, portanto compatíveis com as da literatura, mas apresentam a vantagem de ter distribuições completamente conhecidas.



## 4.7 Estimativas e Comparações entre efeitos médios

Verifica-se através da tabela 5, que para as testemunhas (Achat e Bintje) há maior precisão na estimativa do efeito médio de seus valores genéticos.

Os demais cruzamentos são menos repetidos, apresentando assim, uma amplitude intervalar maior, na maioria dos casos próximo de 1.

Nas Figuras 5B a 9B foram comparadas as densidades a posteriori de cruzamentos em todos os experimentos analisados e verificou-se que, em geral, o erro padrão de todos os efeitos médios foi menor no experimento em que se considerou um número maior de cruzamento (experimento completo).

Conforme se observa nos gráficos, a amplitude das distribuições depende principalmente da precisão das estimativas, que está diretamente relacionada com os diferentes tamanhos dos experimentos.

Deve-se considerar que foram geradas cadeias amostrais para cada parâmetro; e assim, amostras para qualquer combinação linear destes podem ser realizadas de forma a obter as densidades a posteriori de interesse (em especial contrastes entre cruzamentos, se for o caso).

Os quatro melhores cruzamentos, de acordo com suas médias, foram, na ordem decrescente, MCR-16, MCR-18, MCR-17 e Bintje (testemunha).

Entre as duas testemunhas, Bintje e Achat, na ordem decrescente de médias foram classificados os seguintes cruzamentos: MCR-14, MCR-3, MCR-12, MCR-13, MCR-15, MCR-11, MCR-2.

Uma das vantagens da abordagem bayesiana é que ele propicia a construção de intervalos de credibilidade exatos para as estimativas de parâmetros genéticos (Tabela 5).

**TABELA 5: Efeitos de cruzamentos (e testemunhas comerciais) para o modelo linear generalizado misto.**

<b>Cruz.</b>	<b>LI</b>	<b>Média</b>	<b>LS</b>	<b>Amplitude</b>	<b>EP</b>	<b>Erro MC</b>
MCR-16	-0,096	0,453	1,008	1,104	0,277	0,006
MCR-18	-0,067	0,418	0,913	0,980	0,253	0,006
MCR-17	-0,249	0,354	1,002	1,250	0,318	0,006
<b>bintje</b>	<b>0,007</b>	<b>0,321</b>	<b>0,644</b>	<b>0,637</b>	<b>0,165</b>	<b>0,005</b>
MCR-14	-0,286	0,233	0,789	1,074	0,273	0,006
MCR-3	-0,338	0,233	0,735	1,073	0,276	0,006
MCR-12	-0,312	0,185	0,771	1,082	0,275	0,006
MCR-13	-0,351	0,183	0,711	1,062	0,274	0,006
MCR-15	-0,342	0,182	0,739	1,081	0,278	0,006
MCR-11	-0,411	0,125	0,692	1,102	0,279	0,006
MCR-2	-0,465	0,080	0,644	1,109	0,280	0,006
<b>Achat</b>	<b>-0,256</b>	<b>0,079</b>	<b>0,383</b>	<b>0,639</b>	<b>0,164</b>	<b>0,005</b>
MCR-5	-0,473	0,072	0,643	1,115	0,281	0,006
MCR-6	-0,508	0,044	0,567	1,075	0,276	0,006
MCR-9	-0,526	0,028	0,602	1,128	0,286	0,007
MCR-4	-0,509	0,012	0,574	1,083	0,279	0,006
MCR-24	-0,545	-0,004	0,562	1,107	0,282	0,006
MCR-25	-0,612	-0,041	0,492	1,104	0,278	0,006
MCR-1	-0,602	-0,072	0,480	1,082	0,275	0,006
MCR-8	-0,617	-0,101	0,437	1,053	0,272	0,006
MCR-7	-0,641	-0,104	0,422	1,063	0,271	0,006
MCR-23	-0,661	-0,134	0,411	1,071	0,276	0,007
MCR-19	-0,756	-0,234	0,271	1,028	0,260	0,006
MCR-20	-0,783	-0,243	0,314	1,097	0,278	0,005
MCR-10	-0,911	-0,399	0,197	1,108	0,283	0,006
MCR-22	-1,310	-0,724	-0,172	1,138	0,296	0,006
MCR-21	-1,581	-0,976	-0,397	1,184	0,303	0,006

Foram realizados contrastes envolvendo cada um desses cruzamentos e as testemunhas Bintje e Achat com o objetivo de verificar o ganho genético esperado pela seleção. Os contrastes construídos estão apresentados na Tabela 6, como segue:

TABELA 6: Estatísticas descritivas referente a contrastes para alguns cruzamentos.

Contrastes	LI	Média	LS	EP	Erro MC
C16 vs Bintje	-0,4072	0,1317	0,7049	0,2833	0,0047
C18 vs Bintje	-0,4104	0,0971	0,5701	0,2528	0,0041
C17 vs Bintje	-0,5991	0,0334	0,6649	0,3225	0,0054
Contraste*	-0,2973	0,0874	0,4491	0,1929	0,0032
C14 vs Achat	-0,3709	0,1537	0,7153	0,2773	0,0046
C3 vs Achat	-0,4301	0,1533	0,6815	0,2835	0,0042
C12 vs Achat	-0,4558	0,1053	0,6611	0,2803	0,0043
C13 vs Achat	-0,4086	0,1039	0,6901	0,2819	0,0044
C15 vs Achat	-0,4816	0,1024	0,6284	0,2835	0,0043
C11 vs Achat	-0,5155	0,0458	0,6195	0,2882	0,0045
C2 vs Achat	-0,5302	0,0005	0,5928	0,2879	0,0048
Contraste**	-0,2205	0,0950	0,3857	0,1561	0,0023
Contraste***	-0,1438	0,2082	0,5482	0,1767	0,0029
Contraste****	-0,1859	0,0444	0,2767	0,1171	0,0019

\* três melhores vs Bintje;

\*\* 5<sup>o</sup> ao 11<sup>o</sup> melhor vs Achat;

\*\*\* três melhores vs Bintje e Achat;

\*\*\*\*10 melhores vs Bintje e Achat.

Por meio dos valores médios estimados para cada contraste, relacionados na Tabela 6, pode-se calcular o ganho na seleção de um genótipo de um determinado grupo com relação a genótipos selecionado em outro grupo, na escala original do número de tubérculos graúdos. Por exemplo, o ganho em selecionar um genótipo entre os três melhores em relação à testemunha Bintje pode ser de  $e^{0,0874} - e^0 = 0,0913$ , ou seja, 9,13%. Para um genótipo selecionado entre os três melhores com relação a Bintje e Achat, o ganho de

seleção pode ser de  $e^{0,2082} - e^0 = 0,2314$  ou 23,14%. Porém, deve ser ressaltado que há uma grande similaridade entre os tratamentos, pois os intervalos de credibilidade para todos contrastes abrange o valor zero.

#### **4.8 Distribuição dos efeitos médios de cruzamentos e de parcelas**

A Figura 9B e Tabela 5 indicaram valores de efeito médio de cruzamento baixo comparado com os demais, principalmente para os cruzamentos MCR-21 e MCR-22; esses dois valores tendem a não atender a normalidade, mas, de forma geral, os efeitos foram normalmente distribuídos, conforme verificou-se através da Figura 12B.

Através do gráfico qqnormal (percentis amostrais versus percentis esperados de uma distribuição normal padrão) para efeitos médios de parcela, representado na Figura 13B, verificaram-se valores discrepantes para as parcelas p70 e p71, nas quais o número de tubérculos graúdos é pequeno ou nulo para a maioria das plantas. Os efeitos médios obtidos para cada uma foram -1,00276 e -0,50636, nas quais foram dispostos os cruzamentos MCR-21 e MCR-22, respectivamente, com genealogias diferentes conforme Tabela 2.

Os menores efeitos de parcelas (p70 e p71) foram justificados através dos cruzamentos nelas dispostos (MCR-21 e MCR-22), que também apresentaram menores efeitos entre os cruzamentos.

O grupo de parcelas p18, p24, p47, p76, p90 e p104 se destacou quanto ao efeito médio mais elevado (Figura 13B), e os cruzamentos dispostos em cada uma foram Bintje, MCR-24, MCR-4, MCR-15, MCR-6 e MCR-20, que também apresentam genealogia diferente conforme Tabela 2.

**TABELA 7: Número médio de tubérculos graúdos esperados para cada cruzamento e respectivos Intervalos de Probabilidade (95%).**

<b>Cruz</b>	<b>LI</b>	<b>Média</b>	<b>LS</b>
MCR-16	2,8553	4,9440	8,6122
MCR-18	2,9393	4,7739	7,8317
MCR-17	2,4502	4,4780	8,5607
Binje	3,1651	4,3326	5,9845
MCR-14	2,3612	3,9676	6,9184
MCR-3	2,2416	3,9676	6,5547
MCR-12	2,3006	3,7817	6,7949
MCR-13	2,2126	3,7741	6,3992
MCR-15	2,2326	3,7703	6,5809
MCR-11	2,0838	3,5614	6,2788
MCR-2	1,9742	3,4047	5,9845
Achat	2,4331	3,4013	4,6098
MCR-5	1,9585	3,3776	5,9785
MCR-6	1,8911	3,2843	5,5410
MCR-9	1,8574	3,2322	5,7384
MCR-4	1,8892	3,1809	5,5799
MCR-24	1,8224	3,1304	5,5134
MCR-25	1,7043	3,0167	5,1406
MCR-1	1,7215	2,9246	5,0793
MCR-8	1,6958	2,8410	4,8655
MCR-7	1,6556	2,8325	4,7931
MCR-23	1,6228	2,7488	4,7407
MCR-19	1,4758	2,4872	4,1213
MCR-20	1,4365	2,4649	4,3024
MCR-10	1,2639	2,1089	3,8274
MCR-22	0,8480	1,5237	2,6463
MCR-21	0,6467	1,1843	2,1131

Não se verificou relação entre o efeito médio de parcelas para os grupos apresentados acima e parentesco comum para os cruzamentos.

Para efeito médio de parcelas, verificou-se que foram normalmente distribuídos, conforme se observa na Figura 13B, porém algumas parcelas apresentaram valores discrepantes no primeiro e segundo grupos descritos acima.

Na Tabela 7 estão relacionados os valores médios de tubérculos esperados, na escala original, para cada cruzamento obtidos a partir da transformação inversa dada por  $\mu_i = \exp(\eta_i) = 3,1430\exp(c_j)$ , em que  $\exp(c_j)$  é o efeito associado ao cruzamento “j”.

O interessante é notar, no entanto, que a inferência bayesiana permitiria estimar intervalos de credibilidade para o ganho de seleção fazendo, por exemplo, um contraste para comparação dos dez melhores genótipos com relação a todos, ou seja, qualquer contraste de interesse.

## **5 CONCLUSÃO**

- **A aplicação da inferência bayesiana para a obtenção de estimativas de parâmetros através das distribuições a posteriori em experimentos de melhoramento genético constitui uma técnica elegante e confiável.**
- **As distribuições de densidade a posteriori para cada parâmetro permitem a realização de inferências sobre quaisquer parâmetros de interesse usando de estatísticas simples como média, intervalos de credibilidade e outras.**
- **Amostras de combinações lineares ou não lineares de parâmetros constituem uma boa forma de comparar genótipos e obter estimativas de herdabilidades.**

## REFERÊNCIAS BIBLIOGRÁFICAS

AMARO, G. B. Seleção precoce de clones de batata para caracteres do tubérculo. Lavras, 2002. 73 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG.

BARTLETT, M. S. The use of transformations. *Biometrics*, Washington, v. 3, n. 1, p. 39-53, Mar. 1947.

BINK, M. C. A. M.; VAN ARENDONK, J. A. M.; QUAAS, R. L. Bayesian estimation of dispersion parameters with a reduced animal model including polygenic and QTL effects. *Genetics Selection Evolution*, Wageningen, v. 30, n. 2, p 103-125, Mar./Apr. 1998.

BOX, G. E. P.; TIAO, G. C. *Bayesian inference in statistical analysis*. New York: John Wiley & Sons, 1992. 588 p.

CASELLA, G.; GEORGE, E. Explaining the gibbs sampler. *The American Statistician*, Alexandria, v. 46, n. 3, p. 167-174, Aug. 1992.

DINIZ, M. C. D. R. Número de clones por família, seleção clonal e seleção de famílias em programas de melhoramento de batata. 2002. 123 p. Dissertação (Mestrado Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, MG.

GAMERMAN, D. Simulação estocástica via cadeias de markov. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA - SINAPE, 21., 1996, Caxambu. *Anais...* Caxambu, ABE - Associação Brasileira de Estatística, 1996. 196 p.

GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches for calculating marginal densities. *Journal of the American Statistical Association*, London, v. 85, n. 410, p. 398-409, June 1990.

GELMAM, A.; RUBIN, D. B.; CARLIN, J. B.; STERN, H. S. *Bayesian data analysis: texts in Statistical Science*. Chapman & Hall, 1995. 526 p.

GELMAM, A.; RUBIN, D. B. Inference from interative simulation using multiple sequences. *Statistical Science*, Hayward, v. 7, n. 4, p. 457-511, 1992.



**GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, Los Alamitos, v. 6, n. 6, p. 721-741, 1984.**

**HASTINGS, W. K. Monte carlo sampling methods using Markov chains and their applications. Biometrika, London, v. 57, n. 1, 97-109, 1970.**

**HINKELMANN, K.; KEMPTHORNE, O. Design and analysis of experiments, New York: E, Wiley-Interscience, 1994. 495 p.**

**JEFREYS, H. Theory of Probability. Clarendon Press, Oxford. 1961.**

**METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TLLER, E. Equations of state calculations by fast computing machines. Journal of Chemical Phisics, Woodbury, v. 21, n. 6, 1087-1092, 1980.**

**McCULLAGH, P.; NELDER, J. A. Generalized linear model. 2. ed. London: chapaman and Hall, 1989. 511 p.**

**McCULLOCH, C. E.; SEARLE, S. R. Generalized, linear, and mixed models. New York: E. Willey-Interscience, 2001. 324 p.**

**MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. Introduction to the theory of statistics. Tokio. McGraw-Hill. Kogakusha, 3. ed. 1974. 564 p.**

**NATARAJAN, R.; KASS, R. E. Reference bayesian methods for generalized linear mixed models. American Statistical Association, Alexandria, v. 95, n. 449, p. 227-237, Mar. 2000.**

**NELDER, J. A; WENDDERBURN, R. W. M. Generalized linear model. Journal of the Royal Statistical Society A, London, v. 135, n. 3, p. 370-384, 1972.**

**RAFTERY, A. L.; LEWIS, S. comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. Statistical Science, Hayward, v. 7, n. 4, 493 - 497. 1992a**

**RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. Experimentação em genética e melhoramento de plantas. Lavras: UFLA, 2000. 326 p.**

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

SHALL, R. Estimation in generalized linear models with random effects. *Biometrika*, London, v. 78, n. 4, p. 719-727, Dec. 1991.

SMITH, B. J. **Bayesian output analysis program (BOA)**. Disponível em: <[http://www.public-health.uiowa.edu/boa/boa\\_user.pdf](http://www.public-health.uiowa.edu/boa/boa_user.pdf)>. Version 1. 0 User's Manual, 2003. Acesso em: 2003.

SORENSEM, D. **Gibbs sampling in quantitative genetics**. Foulun, 1996. 186 p.

SUN, D.; SPECKMAN, P. L.; TSUTAKAWA, R. K. Random effects in generalized linear mixed models (GLMMs). In: **Generalized linear models: a byesian perspective**. edited by Dey, Ghosh and Mallick, 2000. p. 23-39.

YANG, C. C.; MUTHÉN B. O.; YANG, C. C. Finite mixture multivariate generalized linear models using gibbs sampling and E-M algorithms. **Proceedings National Science Council**, Roca, Washington, v. 23, n. 6, p 695-702, 1999.

YATES, F. Incompletes randomized blocks. *Annals of Eugenics*, London, v. 7, p. 121-140, 1936.

ZEGER, S. L.; KARIM, M. R. Generalized linear models With random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, Alexandria, v. 86, n. 413, p. 79-86, Mar. 1991.

# ANEXOS

<b>ANEXOS</b>	<b>PÁGINA</b>
Anexo A.....	56
Anexo B.....	60
Anexo C.....	73

**ANEXO A Critérios de convergência para as distribuições a posteriori para cada parâmetro.**

**TABELA 1A: Estimativas dos indicadores de convergência para os critérios de Brooks, Gelman & Rubin e Raftery & Lewis.**

Parâmetro	Brooks, Gelman & Rubin		Raftery & Lewis		Erro MC
	R	0,975	Pulo	queima	
$\sigma^2_g$	1,0032	1,0168	2	4	0,0006
$\sigma^2_p$	1,0007	1,0036	1	2	0,0007
$\sigma^2_c$	1,0002	1,0011	1	2	0,0024
$\beta_1$	1,0021	1,0083	3	9	0,0062
$\beta_2$	0,9999	1,0005	1	3	0,0065
$\beta_3$	1,0003	1,0004	2	6	0,0064
$\beta_4$	0,9999	1,0003	2	4	0,0065
$\beta_5$	1,0005	1,0031	2	6	0,0066
$\beta_6$	1,0002	1,0003	2	6	0,0064
$\beta_7$	1,0006	1,0042	2	4	0,0064
$\beta_8$	1,0024	1,0131	4	8	0,0061
$\beta_9$	0,9999	1,0002	2	4	0,0062
$\beta_{10}$	1,0016	1,0044	2	6	0,0064
$\beta_{11}$	1,0015	1,0066	1	3	0,0064
$\beta_{12}$	1,0016	1,0031	2	4	0,0060
$\beta_{13}$	1,0007	1,0047	2	4	0,0063
$\beta_{14}$	1,0002	1,0005	2	6	0,0064
$\beta_{15}$	0,9999	1,0001	1	3	0,0061
p1	1,0010	1,0034	1	2	0,0051
p2	1,0006	1,0037	1	2	0,0042
p3	1,0022	1,0111	1	2	0,0051
p4	0,9998	0,9998	1	2	0,0047
p5	1,0005	1,0022	1	2	0,0044
p6	1,0000	1,0005	1	2	0,0052
p7	1,0004	1,0007	1	2	0,0048
p8	0,9997	0,9998	1	2	0,0052
p9	1,0000	1,0001	1	2	0,0050
p10	1,0004	1,0024	1	2	0,0044
p11	1,0005	1,0009	1	2	0,0045
p12	1,0003	1,0012	1	2	0,0048
p13	0,9999	1,0004	1	2	0,0057
p14	1,0000	1,0008	1	2	0,0053
p15	1,0014	1,0067	1	2	0,0059
p16	0,9998	1,0000	1	3	0,0046
p17	1,0004	1,0029	1	2	0,0052

(continua....)

p18	1,014	1,000	1,062	1	2	0,0048
p19	1,000	1,000	1,012	1	3	0,0050
p20	0,999	0,999	0,999	1	3	0,0048
p21	0,999	1,000	1,000	1	2	0,0054
p22	1,000	1,001	1,001	1	2	0,0055
p23	1,006	1,006	1,037	1	2	0,0049
p24	1,005	1,005	1,018	1	2	0,0048
p25	1,002	1,002	1,006	2	6	0,0047
p26	1,000	1,000	1,007	1	3	0,0048
p27	1,002	1,002	1,002	1	2	0,0057
p28	1,017	1,017	1,097	1	2	0,0058
p29	1,002	1,002	1,018	1	2	0,0044
p30	1,010	1,010	1,025	1	2	0,0057
p31	0,999	1,003	1,003	1	2	0,0052
p32	1,017	1,017	1,093	2	4	0,0044
p33	1,011	1,011	1,059	1	2	0,0056
p34	1,001	1,001	1,015	1	2	0,0052
p35	1,005	1,005	1,034	1	3	0,0052
p36	0,999	1,002	1,002	1	2	0,0050
p37	1,000	1,000	1,007	1	2	0,0048
p38	0,997	0,997	0,997	1	2	0,0045
p39	1,001	1,001	1,015	1	3	0,0045
p40	1,001	1,001	1,015	1	2	0,0048
p41	1,002	1,002	1,014	1	2	0,0047
p42	0,999	1,003	1,003	1	2	0,0044
p43	0,999	1,000	1,000	1	2	0,0056
p44	0,998	1,000	1,000	1	2	0,0055
p45	1,021	1,008	1,008	1	2	0,0053
p46	0,999	1,003	1,003	1	2	0,0050
p47	1,008	1,008	1,036	1	2	0,0054
p48	1,004	1,004	1,024	1	2	0,0044
p49	1,008	1,008	1,049	1	3	0,0049
p50	1,006	1,006	1,029	1	3	0,0041
p51	1,006	1,006	1,035	1	3	0,0058
p52	1,027	1,027	1,042	1	2	0,0044
p53	0,998	0,998	0,998	1	2	0,0051
p54	1,019	1,022	1,022	1	2	0,0054
p55	1,004	1,017	1,017	1	2	0,0060
p56	1,001	1,013	1,013	1	2	0,0045
p57	1,007	1,043	1,043	1	2	0,0052
p58	1,000	1,003	1,003	1	2	0,0046
p59	1,025	1,085	1,085	1	2	0,0051
p60	1,002	1,003	1,003	1	2	0,0050
p61	1,012	1,057	1,057	1	2	0,0050

(continua...)

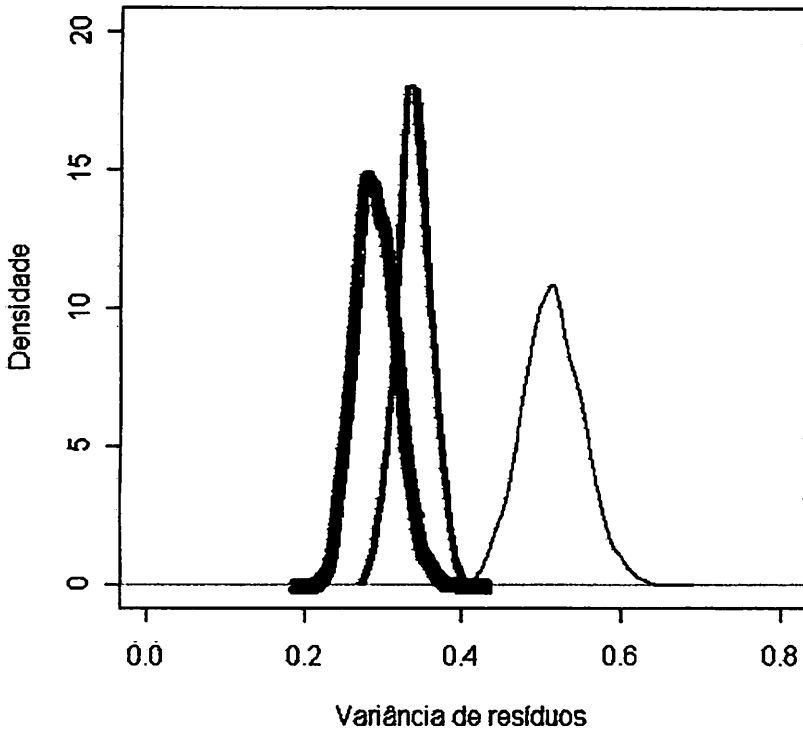
p62	1,003	1,025	1	3	0,0049
p63	1,004	1,013	1	2	0,0044
p64	1,009	1,023	1	2	0,0054
p65	1,006	1,026	1	2	0,0046
p66	0,9999	0,9999	1	2	0,0045
p67	1,006	1,0032	1	2	0,0045
p68	1,001	1,0013	1	2	0,0046
p69	0,9998	1,0000	1	2	0,0051
p70	1,0007	1,0043	1	2	0,0059
p71	1,011	1,0048	1	2	0,0064
p72	1,0003	1,0023	1	2	0,0049
p73	0,9999	1,0005	1	2	0,0043
p74	0,9999	0,9999	1	2	0,0047
p75	1,0020	1,0110	1	3	0,0051
p76	1,0002	1,0016	1	2	0,0044
p77	1,0010	1,0055	1	2	0,0050
p78	1,0007	1,0038	1	2	0,0049
p79	1,0005	1,0032	1	3	0,0049
p80	1,0005	1,0032	1	2	0,0043
p81	0,9998	1,0000	1	3	0,0045
p82	1,0002	1,0016	1	2	0,0047
p83	0,9998	0,9999	1	2	0,0049
p84	0,9997	0,9998	1	2	0,0049
p85	1,0009	1,0054	1	2	0,0051
p86	1,0000	1,0004	1	2	0,0049
p87	0,9999	1,0002	1	3	0,0044
p88	1,0015	1,0082	1	2	0,0044
p89	0,9999	1,0003	1	2	0,0050
p90	1,0001	1,0009	1	2	0,0049
p91	1,0009	1,0037	1	3	0,0045
p92	0,9998	1,0002	1	2	0,0052
p93	0,9997	0,9998	1	2	0,0045
p94	1,0002	1,0012	1	2	0,0055
p95	1,0005	1,0022	1	2	0,0051
p96	0,9998	1,0002	1	2	0,0059
p97	1,0003	1,0027	1	3	0,0054
p98	0,9998	0,9998	1	2	0,0050
p99	0,9998	1,0002	1	3	0,0044
p100	1,0004	1,0028	1	2	0,0051
p101	1,0009	1,0049	1	2	0,0053
p102	1,0011	1,0049	1	2	0,0052
p103	0,9998	0,9998	1	3	0,0046
p104	0,9998	0,9998	1	2	0,0050
p105	1,0001	1,0002	1	2	0,0053

(continua...)

TABELA. Cont.

C1	1,0013	1,0067	1	2	0,0057
C2	1,0009	1,0020	2	2	0,0061
C3	0,9998	0,9998	1	2	0,0058
C4	1,0006	1,0037	1	2	0,0059
C5	1,0000	1,0012	1	2	0,0062
C6	1,0011	1,0058	1	3	0,0059
C7	1,0012	1,0070	1	3	0,0059
C8	0,9999	1,0004	1	3	0,0056
C9	1,0003	1,0004	1	2	0,0066
C10	0,9998	0,9999	1	3	0,0058
C11	1,0004	1,0016	1	2	0,0060
C12	1,0008	1,0038	1	3	0,0061
C13	1,0000	1,0004	1	2	0,0055
C14	1,0014	1,0067	1	2	0,0063
C15	1,0020	1,0104	1	3	0,0060
C16	1,0016	1,0057	1	2	0,0061
C17	0,9999	0,9999	1	2	0,0064
C18	1,0005	1,0006	1	2	0,0056
C19	1,0007	1,0029	1	2	0,0058
C20	1,0006	1,0040	1	2	0,0055
C21	1,0006	1,0041	1	3	0,0064
C22	1,0012	1,0034	1	3	0,0063
C23	0,9999	1,0007	1	2	0,0065
C24	0,9998	0,9998	1	3	0,0057
C25	1,0010	1,0035	1	2	0,0059
C26	0,9999	1,0001	2	6	0,0050
C27	0,9999	1,0004	2	8	0,0051

**ANEXO B Figuras das densidades a posteriori para os diversos parâmetros em cada situação experimental (Componentes das variâncias, efeitos), contrastes e qqnormal para cruzamentos e parcelas.**



**Figura 1B: Densidades de variâncias para resíduos.**

- Experimento completo
- Exp. com exceção das testemunhas
- Exp. considerando apenas testemunhas



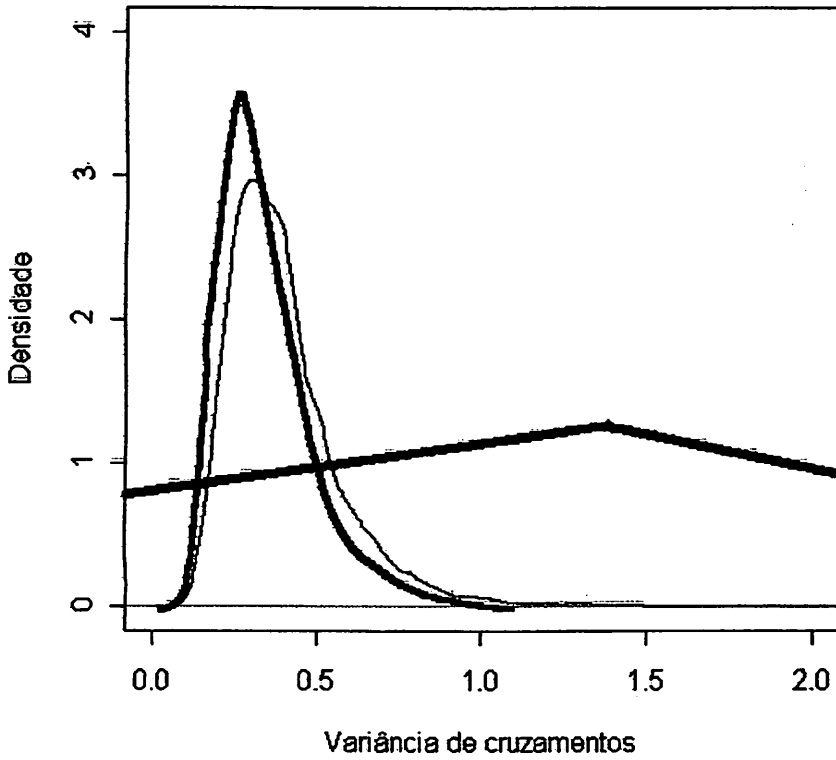


Figura 2B: Densidades de variâncias para cruzamentos.

- Experimento completo
- Exp. com exceção das testemunhas
- Exp. considerando apenas testemunhas

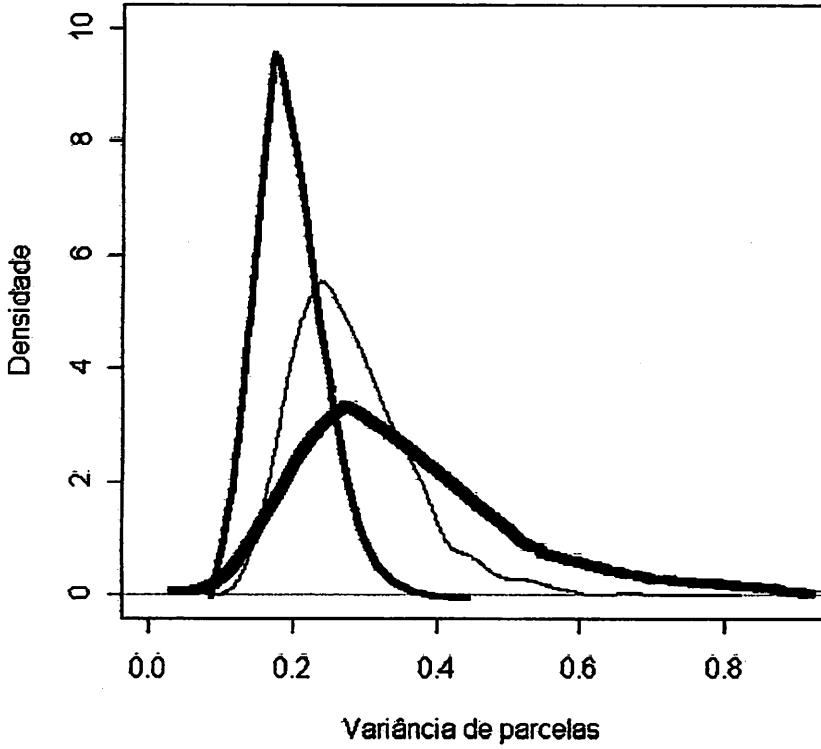


Figura 3B: Densidades de variâncias para parcelas.

- Experimento completo
- Exp. com exceção das testemunhas
- Exp. considerando apenas testemunhas

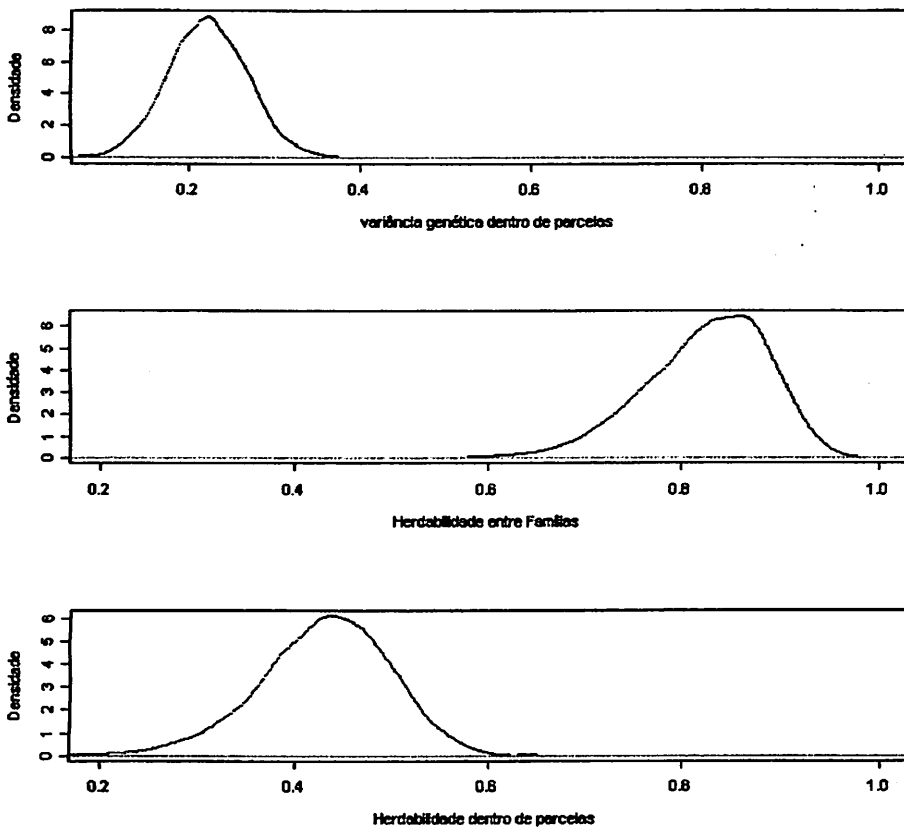


Figura 4B: Densidade a posteriori para variância genética dentro de parcelas, herdabilidade entre famílias e herdabilidade dentro de parcelas.

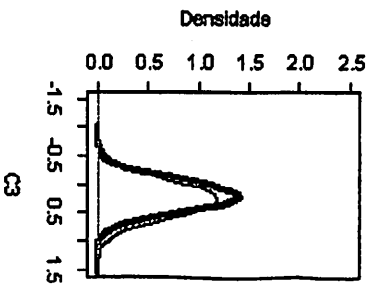
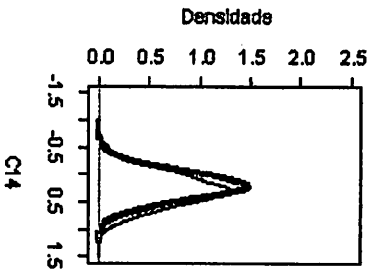
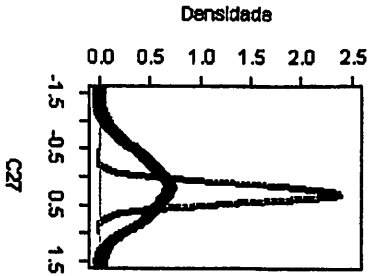
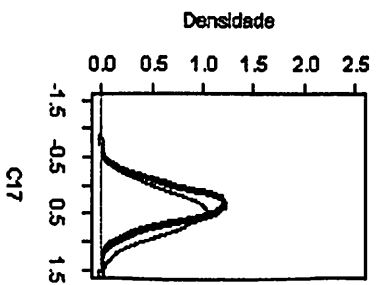
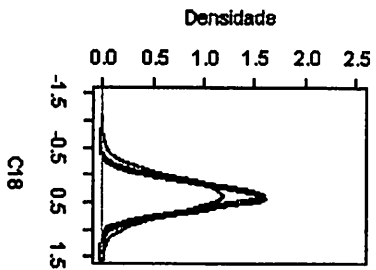
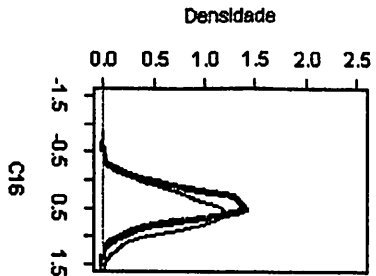


Figura 5B: Densidades dos valores preditos para os cruzamentos na ordem decrescente de médias.  
— Experimento completo  
- - - Exp. com exceção das testemunhas  
— Exp. considerando apenas testemunhas

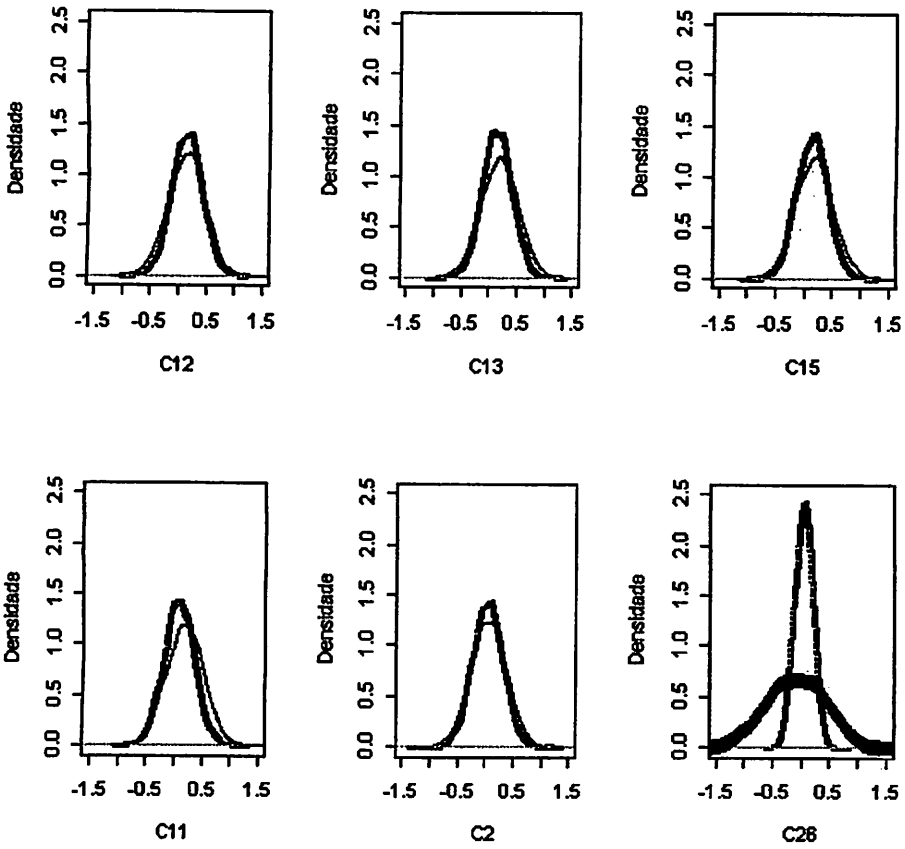


Figura 6B: Densidades dos valores preditos para os cruzamentos na ordem decrescente de médias.

- Experimento completo
- Exp. com exceção das testemunhas
- Exp. considerando apenas testemunhas

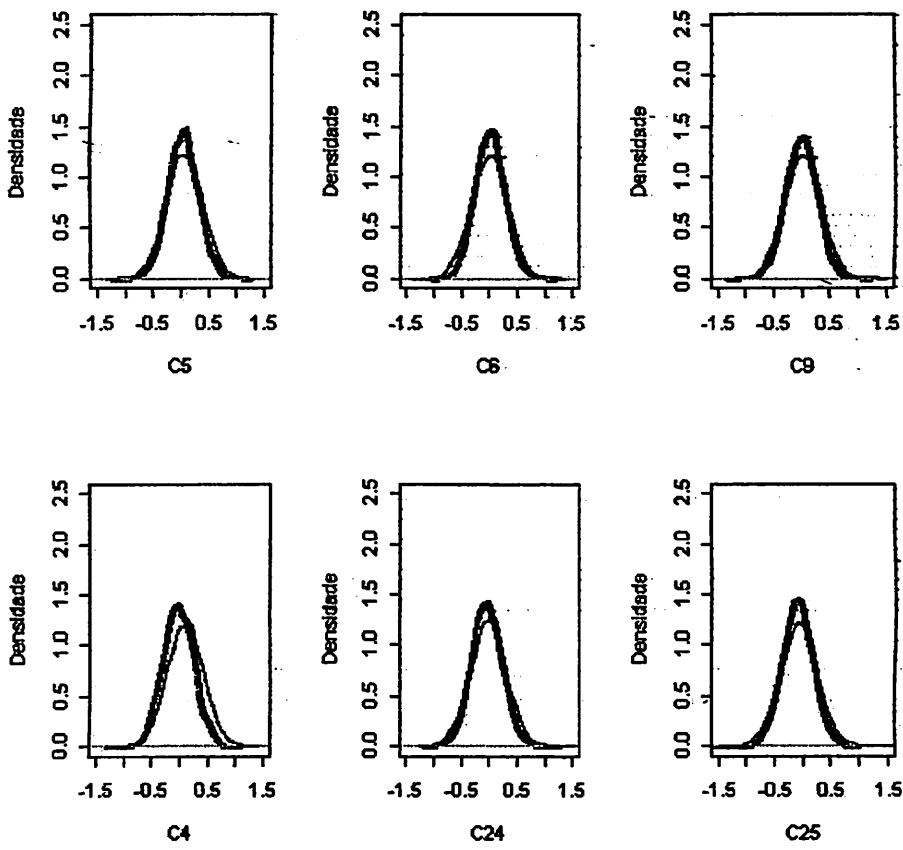


Figura 7B: Densidades dos valores previstos para os cruzamentos na ordem decrescente de médias.

— Experimento completo  
 - - - Exp. com exceção das testemunhas

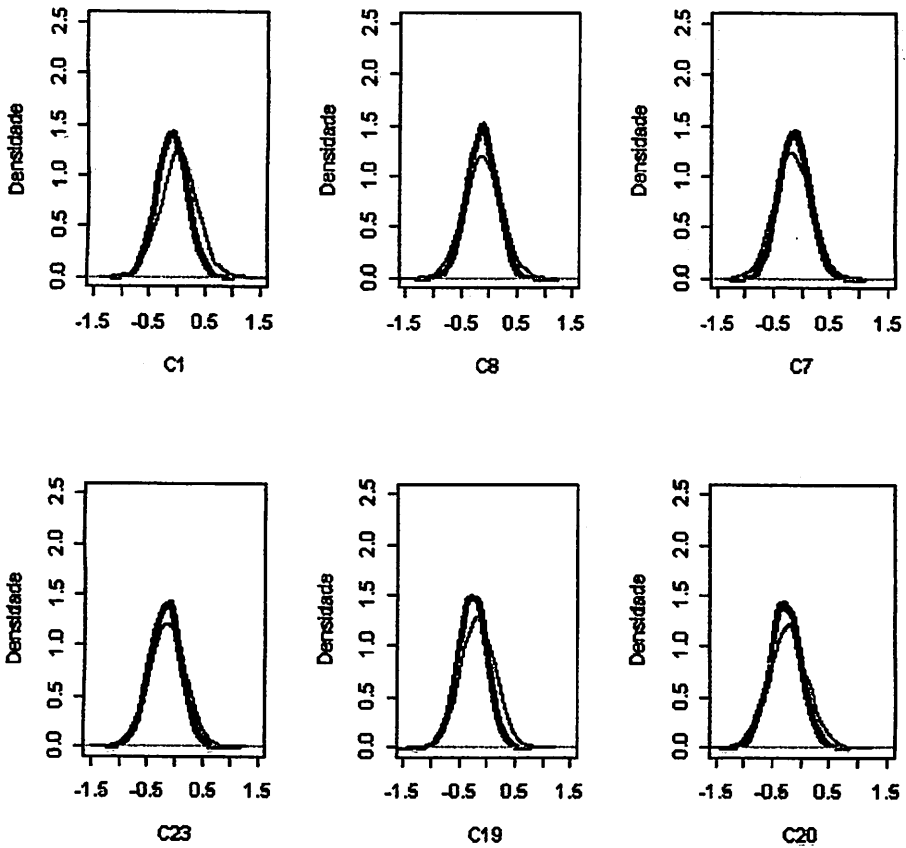


Figura 8B: Densidades dos valores preditos para os cruzamentos na ordem decrescente de médias.

— Experimento completo  
 - - - Exp. com exceção das testemunhas

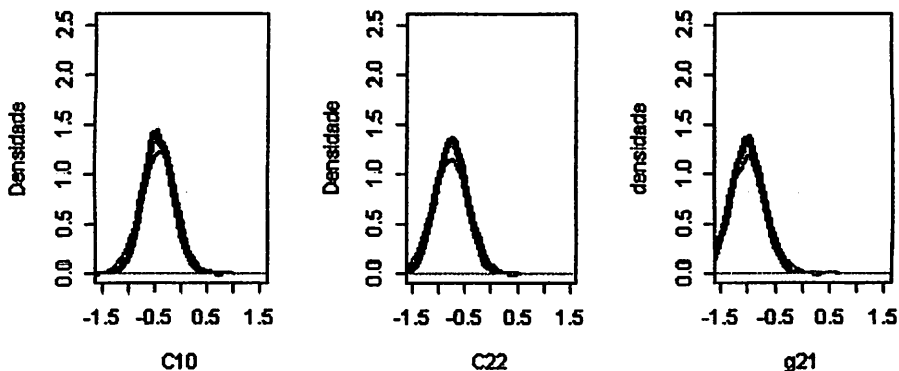


Figura 9B: Densidades dos valores preditos para os cruzamentos 10, 22 e 21 na ordem decrescente de médias.

— Experimento completo  
 - - - Exp. com exceção das testemunhas

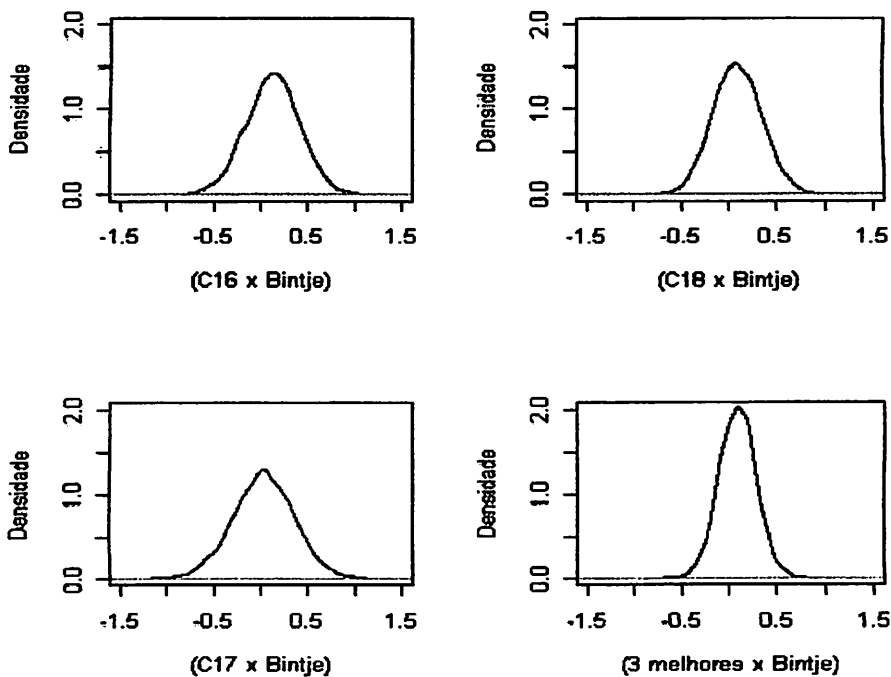


Figura 10B: Gráficos das densidades a posteriori para contrastes.



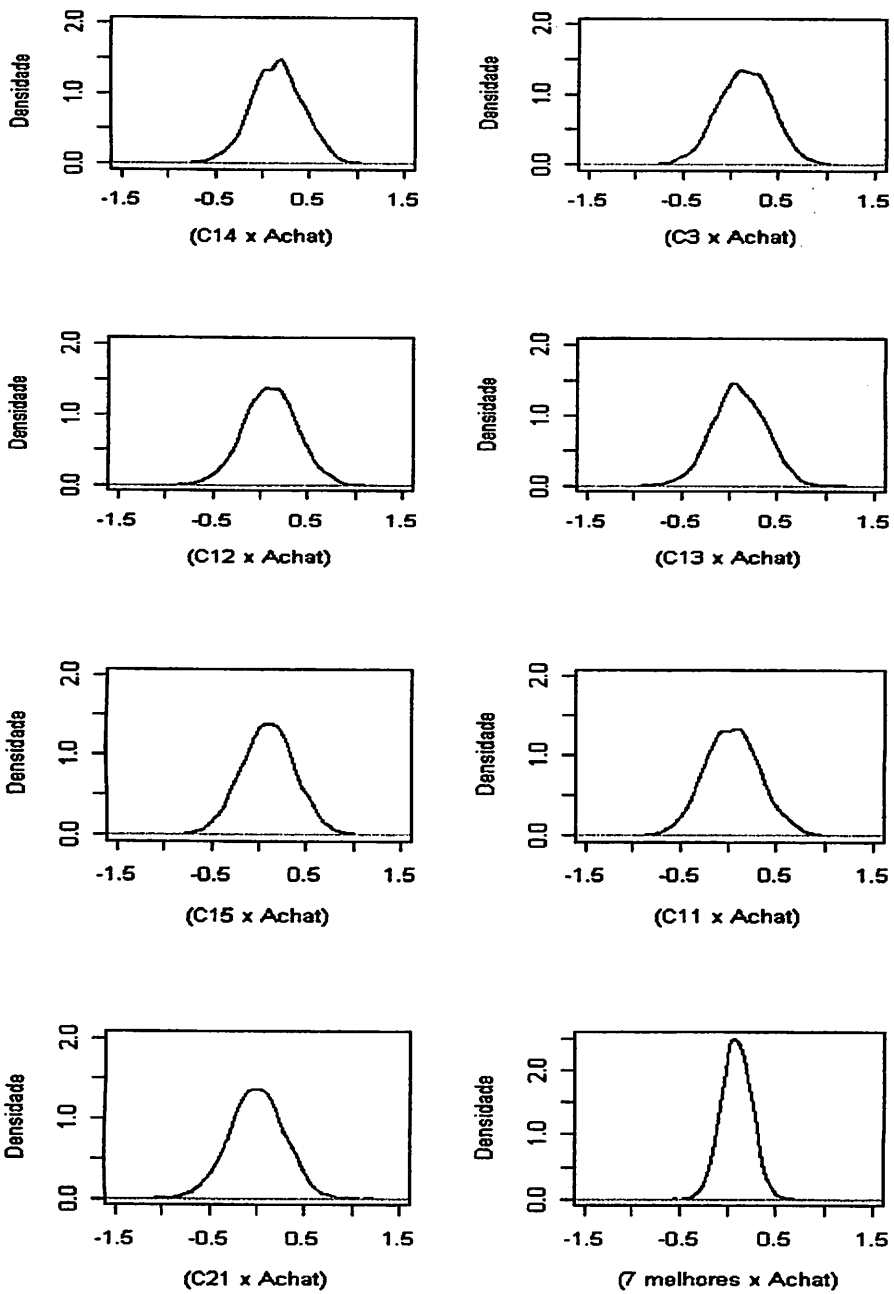
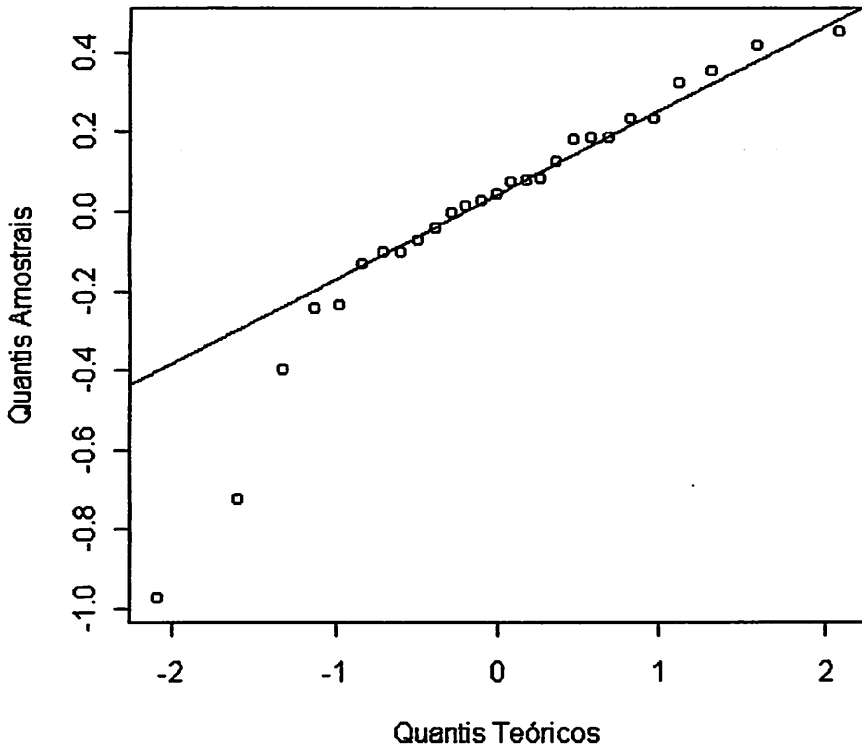


Figura 11B: Gráficos das densidades a posteriori para contrastes.



**Figura 12B:** Gráficos do tipo qqnormal para os efeitos médios de cruzamentos. A cultivar Bintje é a quarta de cima para baixo (da direita para a esquerda) e a cultivar Achat é a décima segunda.

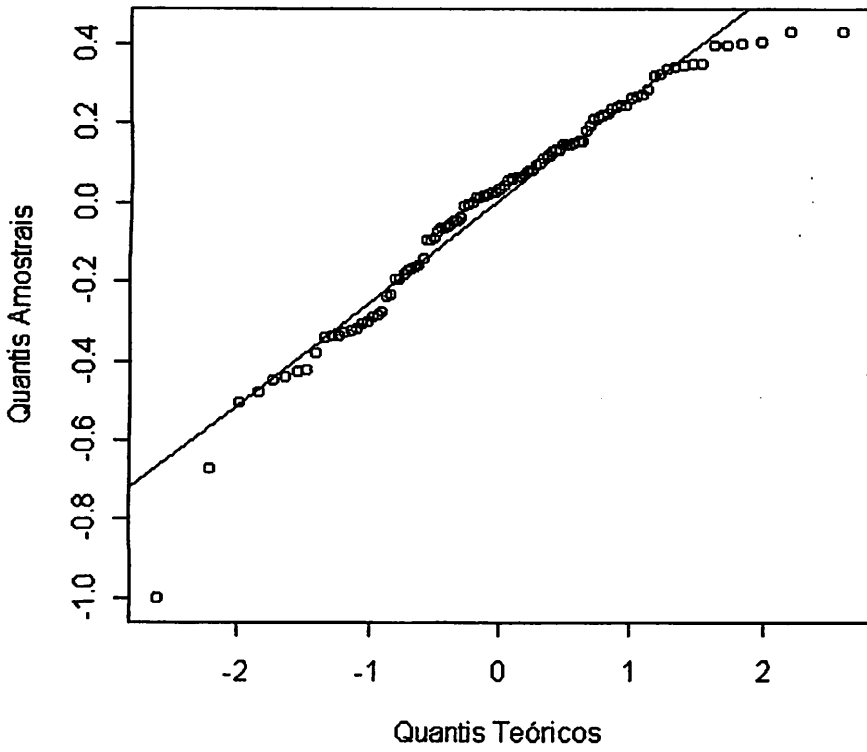


Figura 13B: Gráficos do tipo qqnormal para os efeitos médios de parcelas.

## **ANEXO C Rotina para a implementação da Amostragem Gibbs das distribuições condicionais completas a posteriori.**

```
library(MASS)
dados<- read.table("arquivo de dados",h=T)
attach(dados)

p <- factor(parc)
bl <- factor(bloco)
c <- factor(cruz)
Ng <- Ng
eta <- log(Ng+0.01)

n <- length(Ng)

Xb <- matrix(0,n,15)
Xp <- matrix(0,n,105)
Xc <- matrix(0,n,27)

for (i in 1:n)
{
  Xb[i,bloco[i]] <- 1
  Xp[i,parc[i]] <- 1
  Xc[i,cruz[i]] <- 1
}

Z<-cbind(Xp,Xc)

# Cadeia de markov para componentes da variância e #
# efeitos fixos e aleatórios e eta #

# Inferência Bayesiana #

iter <- 100000
SS <- 3800
BI <- 5000
TH <- 25
```

```

amostra <- matrix(0,SS,(3+15+105+27))
feta0 <- matrix(0,n,1)
geta0 <- matrix(0,n,1)

# Pontos iniciais ( Chutes) #

vc <- 1
vp <- 1
ve <- 10

# Priors

rc <- 5
sc <- 1

rp <- 5
sp <- 1

re <- 5
se <- 10

# Delineamento #

Q <- cbind(Xb,Z)
theta <- ginv(t(Q)%*%Q) %*% t(Q)%*% eta

beta <- theta[1:15]
u <- theta[16:147]

e <- eta-Q%*%theta
vu <- as.real((t(u)%*%u)/2)
ve <- as.real((t(e)%*%e)/2)

F <- ginv(t(Xb)%*%Xb)
cF <- chol(F)

```

```

# Loop MCMC #

for (i in 1:iter)
{
ap <- as.real(ve/vp)
ac <- as.real(ve/vc)

S <- rbind(cbind(ap*diag(105),matrix(0,105,27)),
           cbind(matrix(0,27,105),ac*diag(27)))

# Condicional para efeitos fixos #

meanf <- ginv(t(Xb)%*%Xb)%*%t(Xb)%*%(eta-Z %*% u)
kf <- morm(15)
beta <- sqrt(ve) * t(cF) %*% kf + meanf

# Condicional para efeitos aleatórios #

M <- ginv(t(Z)%*%Z + S)

meanr <- M%*%t(Z)%*%(eta-Xb%*%betah)
kr <- morm(132)
sr <- sqrt(ve) * chol(M)
u <- t(sr) %*% kr + meanr

# Condicional para variância residual #

e <- eta - Xb%*%betah-Z%*%u
c1 <- (n+re)/2
c2 <- 2/(t(e)%*%e+re*se)
ve <- c2*rchisq(1,c1)
ve <- as.real(1/ve)

# Condicional para variância de parcelas #

parcela <- u[1:105]
c1 <- (105+rp)/2
c2 <- 2/(t(parcela)%*%parcela+rp*sp)
vp <- c2*rchisq(1,c1)
vp <- as.real(1/vp)

```

```
# Condicional para variância de cruzamentos #
```

```
cruzamento <- u[106:132]  
c1          <- (27+rc)/2  
c2          <- 2/(t(cruzamento)%*%cruzamento+rc*sc)  
vc          <- c2*rchisq(1,c1)  
vc          <- as.real(1/vc)
```

```
# Condicional para eta #
```

```
# Amostragem para eta #
```

```
meanmh1 <- Xb %*% betah + Z %*% u  
sdrmh1  <- rep(sqrt(ve),n)  
kmh1    <- morm(n)
```

```
eta0    <- sdrmh1 * kmh1 + meanmh1  
teste   <- runif(n)
```

```
for(j in 1:n)
```

```
{  
  feta0[j] <- exp(-((eta0[j]-meanmh1[j])^2)/(2*ve))  
  geta0[j] <- exp(Ng[j]*eta0[j]-exp(eta0[j])-log(gamma(Ng[j]+1)))  
  if (teste[j] < (feta0[j]*geta0[j]/feta0[j]))  
  {  
    eta[j] <- eta0[j]  
  }  
}
```

```
for(j in 1:SS)
```

```
{  
  if (round((i-BI-1)/TH) = j)  
  {  
    amostra[j,] <- cbind(ve, vp, vc, t(beta), t(parcela), t(cruzamento))  
  }  
}
```