

VINÍCIUS RIBEIRO FARIA

**MELHOR PREDIÇÃO LINEAR IMPARCIAL (BLUP) NO MELHORAMENTO
VEGETAL: SELEÇÃO ENTRE FAMÍLIAS DE MEIOS-IRMÃOS**

Dissertação apresentada à
Universidade Federal de Viçosa, como
parte das exigências do Programa de
Pós-Graduação em Genética e
Melhoramento, para obtenção do título de
“*Magister Scientiae*”.

VIÇOSA
MINAS GERAIS – BRASIL
2008

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

T

F224m
2008

Faria, Vinícius Ribeiro, 1983-

Melhor predição linear imparcial (BLUP) no melhoramento vegetal : seleção entre famílias de meio-irmãos / Vinícius Pinheiro Faria. – Viçosa, MG, 2008. ix, 57f. : il. ; 29 cm.

Inclui anexos.

Orientador: José Marcelo Soriano Viana.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 47-52.

1. Plantas - Melhoramento genético. 2. Melhoramento genético vegetal. 3. BLUP. I. Universidade Federal de Viçosa. II. Título.

CDD 22 ed. 631.52

VINÍCIUS RIBEIRO FARIA

**MELHOR PREDIÇÃO LINEAR IMPARCIAL (BLUP) NO MELHORAMENTO
VEGETAL: SELEÇÃO ENTRE FAMÍLIAS DE MEIOS-IRMÃOS**

Dissertação apresentada à
Universidade Federal de Viçosa, como
parte das exigências do Programa de
Pós-Graduação em Genética e
Melhoramento, para obtenção do título de
“*Magister Scientiae*”.

APROVADA: 24 de julho de 2008.

Prof. Fabyano Fonseca e Silva
(Co-orientador)

Prof. Cosme Damião Cruz
(Co-orientador)

Dr. Willian Silva Barros

Prof. Glauco Vieira Miranda

Prof. José Marcelo Soriano Viana
(Orientador)

À minha família.
Aos meus pais Jandir e Célia.

AGRADECIMENTOS

À Universidade Federal de Viçosa, pela oportunidade de realizar este curso.

À FAPEMIG, pelo auxílio financeiro.

Ao meu orientador, professor José Marcelo Soriano Viana, pela segura orientação, pelo apoio, pela confiança e pela amizade.

Ao professor Fabyano, pelas valiosas sugestões, pela atenção, pelo apoio e pela amizade.

Ao Dr. Willian, pelas ajudas e pela amizade desde os tempos de alojamento.

Aos professores Cosme Damião Cruz e Glauco Vieira Miranda, pelas apreciações e pelas valiosas sugestões.

Aos amigos Aloisio, Aurinelza, Emmanuel, Mauro, João, Fábio, Tassiano, Magno, Ísis, Ramon, Keny, Admilson, Rodrigo e Gabriel, pela colaboração e pela agradável convivência.

A todos os amigos, com os quais convivi desde os tempos de graduação, e aos tantos companheiros de república.

Aos companheiros do futebol, pelas agradáveis manhãs de sábado na Rua Nova.

Aos companheiros da Vaca Magra, Bacana, Batata, Fabrício, Pudim, Leo e seus freqüentadores, Dudim, Dimirso, Primo, Grilo, Marjorie, Michelle, Jorginho da Mercearia, Paulim, Gabriel Sobreira, 2mi2, Claudinei, Monique, Eulália.

Às funcionários Vicente, Márcio e Antônio, pelos esforços despendidos nos trabalhos de campo.

Às funcionárias da secretaria, Rose e Rita, pela atenção e pela amizade.

Ao meu Irmão, Héder, que me trouxe para Viçosa e que sem o qual, seria muito mais difícil estar onde estou hoje.

A todos os que colaboraram para realização deste trabalho, minha sincera gratidão.

BIOGRAFIA

VINÍCIUS RIBEIRO FARIA, filho de Jandir de Faria e Célia Ribeiro de Faria, nasceu na cidade de Dores do Indaiá - MG, em 22 de outubro de 1983.

Realizou os estudos do primeiro e segundo graus, na Escola Estadual Benjamim Guimarães e na Escola Estadual Francisco Campos, respectivamente, até o ano de 2001.

Ingressou no curso de Agronomia no ano de 2002 na Universidade Federal de Viçosa.

Em outubro de 2006, obteve o diploma de Engenheiro Agrônomo pela Universidade Federal de Viçosa.

Em 2006 iniciou, o Programa de Mestrado em Genética e Melhoramento, pela mesma instituição, com ênfase em Genética Quantitativa e Melhoramento de Milho-Pipoca, submetendo-se à defesa de tese em julho de 2008.

SUMÁRIO

RESUMO	vi
ABSTRACT	viii
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	4
2.1. Metodologia de modelos mistos	4
2.2. Estimação de componentes de variância	6
2.3. BLUP no melhoramento genético	11
3. MATERIAL E MÉTODOS	17
3.1. Metodologia de modelos mistos	17
3.1.1. Análises computacionais utilizando modelos mistos	21
3.2. Método dos mínimos quadrados	25
3.3. Ganhos de seleção	28
3.4. Aplicação a dados experimentais	29
4. RESULTADOS E DISCUSSÃO	31
5. RESUMO E CONCLUSÕES	45
REFERÊNCIAS BIBLIOGRÁFICAS	47
ANEXOS	53

RESUMO

FARIA, Vinícius Ribeiro, M.Sc., Universidade Federal de Viçosa, julho de 2008. **Melhor Predição Linear Imparcial (BLUP) no melhoramento vegetal: seleção entre famílias de meios-irmãos.** Orientador: José Marcelo Soriano Viana. Co-orientadores: Fabyano Fonseca e Silva e Cosme Damião Cruz.

O presente trabalho teve como objetivos implementar a metodologia REML/BLUP no melhoramento populacional recorrente, empregando famílias de meios-irmãos, comparar essa metodologia com a análise tradicional por mínimos quadrados ordinário e apresentar as rotinas do SAS para as análises de modelos mistos. O modelo empregado corresponde ao modelo de reprodutor utilizado no melhoramento animal. As análises computacionais foram realizadas utilizando o aplicativo computacional SAS[®] por meio dos seus procedimentos *proc mixed*, *proc inbreed*, *proc iml* e *proc glm*. Como suporte à teoria, foram utilizados dados de três ciclos de seleção recorrente na população de milho-pipoca Viçosa, do Programa de Melhoramento do Setor de Genética, do Departamento de Biologia Geral da Universidade Federal de Viçosa. Os experimentos foram delineados em látice 14 x 14 simples. Em cada parcela dos experimentos, foram avaliadas a produção e a capacidade de expansão (CE) em ml/g. Para cada ciclo de melhoramento, foram estimados os componentes de variância e preditas as médias genóticas, de acordo com a metodologia REML/BLUP. Para cada ciclo também foi feita análise por Mínimos Quadrados Ordinário (OLS), com estimação dos componentes de variância e das médias genóticas. Foram calculados, ainda, os ganhos devidos à seleção em cada ciclo. Também foi verificada a ordenação dos indivíduos selecionados e construídos os intervalos de confiança para os valores genéticos preditos e estimados. Pelas análises de variância, pôde-se verificar que à medida que os ciclos de melhoramento avançaram, houve uma redução significativa na variabilidade para CE, sendo que no terceiro ciclo de seleção não foi detectada variabilidade. Essa redução da variabilidade pode ser explicada pela alta intensidade de seleção, tanto que as 20 famílias selecionadas no terceiro

ciclo eram originárias de apenas nove plantas do primeiro ciclo de seleção. Quanto aos parâmetros genéticos estimados pelas duas metodologias, observou-se que, tanto pela metodologia REML quanto pela OLS, os valores estimados para os componentes de variância foram semelhantes. Para as médias genotípicas, quando não existia parentesco entre os indivíduos, houve concordância entre os melhores indivíduos estimados e preditos, porém, os valores preditos apresentaram intervalos de confiança mais estreitos, indicando que esses valores são mais precisos que os estimados. Ao incluir as informações de parentesco, houve uma redução nos valores de coincidência entre os selecionados e os da correlação de Spearman, indicando que o uso das médias genotípicas preditas na seleção dos indivíduos superiores proporcionará maiores ganhos devidos à seleção.

ABSTRACT

FARIA, Vinícius Ribeiro, M.Sc., Universidade Federal de Viçosa, July, 2008.
Best Linear Unbiased Prediction (BLUP) in plant breeding: selection among half-sib families. Adviser: José Marcelo Soriano Viana. Co-Advisers: Fabyano Fonseca e Silva and Cosme Damião Cruz.

The objectives of the present work were to install the REML/BLUP methodology in the recurrent populational improvement, using half-brothers families, to compare this methodology with the traditional analysis of the ordinary least squares and to present the SAS routines for the analyses of mixed models. The model used corresponds to the beeder model used in the animal improvement. The computer analyses were carried out using the computer software SAS[®] through *proc mixed*, *proc inbreed*, *proc iml* and *proc gml* procedures. As the theory support, data from three cycles of recurrent selection in the population of Voçosa por-corn, of the Programa de Melhoramento do Setor de Genética (Improvement Program of the Genetic Sector), do Departamento de biologia Geral da Universidade Federal de Viçosa (Department of General Biology of the Viçosa Federal University) were used. The experiments were designed in single 14 x 14 lattice. In each block of the experiment, the production and the expansion ability (CE) in ml/g were evaluated. For each improvement cycle, the variance components were estimated the genotypic means were predicted, according to the REML/BLUP methodology. For each cycle the analysis by the ordinary Least Squares (OLS) was also made with the estimation of the variance components and of the genotypic means. Besides, the gains due to the selection in each cycle were calculated. Moreover the ordination of the selected individuals was verified and the confidence intervals for the genetic values predicted and estimated were built. By the variance analysis, it could be verified that as the improvement cycles progressed, there was a significant reduction in the variability for CE, and in the third selection cycle the variability was no more detected. This variability reduction can be explained by the high selection intensity, because the 20 families selected in the third cycle were from only nine plants of the first cycle of selection. Referring the

genetic parameters estimated by the two methodologies, it was observed that both with the REML and the OLS methodology the estimate values for the variance components were similar. For the genotypic means, when there was no relationship among the individuals there was an accordance among the best individuals estimated and predicted, but the predicted values presented narrower confidence intervals, indicating that these values are more accurate than the estimated ones. When the relationship information was included, there was a reduction on the coincidence values between the selected ones and those of the Spearman co-relation indicating that the use of the predicted genotypic means in the selection of superior individuals will provide greater gains due to the selection.

1. INTRODUÇÃO

A predição de variáveis aleatórias e a estimação de componentes de variância e de combinações lineares estimáveis de efeitos fixos e, ou, aleatórios, via metodologia de modelos mistos, apresentam grande relevância no melhoramento genético. No melhoramento animal, essa metodologia já é amplamente utilizada na avaliação genética. Nos últimos anos, com o surgimento de novos aparatos computacionais, vem surgindo grande interesse por parte dos melhoristas de espécies vegetais em trabalhar com essa metodologia. Estudos de Bernardo (1994,1995,1996) indicam que os métodos de estimação/predição via modelos lineares mistos podem ser também aplicados a ensaios genéticos dos programas de melhoramento de milho. Entretanto, esses estudos se baseiam em análises dialélicas e na predição de híbridos não realizados, não havendo trabalhos com modelos mistos aplicados à seleção recorrente.

No melhoramento genético intrapopulacional, o objetivo maior é selecionar os genitores que apresentem os melhores valores genéticos aditivos. Com isso, a eficiência da seleção, admitindo uma adequada experimentação, depende diretamente do emprego de métodos estatísticos que aproveitem ao máximo as informações disponíveis. Vários métodos têm sido propostos no contexto de modelos fixos e mistos, como o método dos quadrados mínimos e métodos baseados na verossimilhança, resultando em diferentes funções dos valores fenotípicos, com a finalidade de predizer os valores genéticos dos indivíduos, como o BLUE (melhor estimador linear não tendencioso), o BLP (melhor preditor linear) e mais recentemente o BLUP (melhor preditor linear não tendencioso) (WHITE e HODGE, 1989).

Na determinação dos valores genéticos, estes mesmos valores podem ser considerados efeitos fixos a serem estimados ou variáveis aleatórias a serem preditas (SEARLE *et al.*, 1992; WHITE e HODGE, 1989). No caso de os valores genéticos serem considerados efeitos fixos, o processo de estimação é feito pelo método dos quadrados mínimos ordinário (OLS), no caso de homogeneidade de variâncias, ou pelo método dos quadrados mínimos ponderado ou pelo método dos quadrados mínimos generalizado, no caso de heterogeneidade de variâncias e não

independência dos erros. RESENDE *et al.* (1996) relataram que os métodos OLS superestimam os valores genéticos, devido ao fato de não considerarem a herdabilidade, além de serem inflacionados pelos efeitos fixos do modelo.

Tratando os valores genéticos como variáveis aleatórias, a predição pode ser efetuada pelo menos em três situações distintas. Entretanto, maior destaque tem sido dado aos preditores BLUP (melhores preditores lineares imparciais) os quais têm sido provados superiores aos demais procedimentos, por serem mais completos e conduzirem à maximização do ganho genético por ciclo de seleção (RESENDE, 1997).

O conhecimento prévio dos componentes de variância e de covariância é necessário na predição dos valores genéticos, quando se faz uso de métodos de predição como o BLUP. Entretanto, estes componentes não são, na maioria das vezes, conhecidos e podem ser estimados por alguma metodologia. Geralmente os componentes de variância são estimados pelo emprego do método OLS em esquemas de cruzamentos, como dialelos e na análise de variância de experimentos de avaliação de famílias. Entretanto, este método não apresenta características ótimas em situação de dados não balanceados, requer suposições menos realistas e não considera as covariâncias existentes entre os indivíduos avaliados (HENDERSON, 1984). Outra grande desvantagem dos métodos de mínimos quadrados é a possibilidade de obtenção de estimativas negativas de componentes de variância. Uma alternativa ao uso de OLS como método para estimar componentes de variância, bem como predizer valores genéticos em ensaios, é o uso da teoria de modelos mistos. Nesta abordagem, estimam-se os componentes por Máxima Verossimilhança Restrita.

Atualmente, vários softwares estão disponíveis para a análise de dados experimentais utilizando modelos mistos. Entre os mais utilizados para o melhoramento animal se destaca o MTDFREML - Multiple Trait Restricted Maximum Likelihood, desenvolvido por Boldman *et al.* (1993). No melhoramento florestal, merece destaque o software Selegen REML/BLUP (RESENDE, 2002). Uma outra opção para as análises, que permite a

estimação dos efeitos fixos e aleatórios de forma prática e flexível, é empregar o software SAS, utilizando o procedimento *proc mixed*.

Com base no exposto, o presente trabalho teve como objetivos a implementação da metodologia REML/BLUP no melhoramento genético empregando famílias de meios-irmãos, comparar essa metodologia aos resultados obtidos com análise tradicional de mínimos quadrados e apresentar as rotinas do SAS para as análises de modelos mistos.

2. REVISÃO DE LITERATURA

2.1. Metodologia de modelos misto

Um modelo estatístico tem como função explicar as observações de uma variável dependente por meio dos efeitos diferenciais que se atribuem a uma outra série de variáveis independentes. Tais efeitos podem se classificar em fixos ou aleatórios. Quando a um determinado fator é atribuído um conjunto finito de tratamentos ou níveis, tal fator é considerado de efeito fixo (SEARLE *et al.*, 1992). Nesse caso, as conclusões são restritas aos tratamentos estudados. Por outro lado, quando a um fator se atribui um conjunto infinito de tratamentos, tal fator é considerado de efeito aleatório, e as inferências são feitas sobre as populações das quais os tratamentos foram amostrados. Com base na natureza fixa ou aleatória dos efeitos contidos em um modelo estatístico, este pode se classificar em fixo, aleatório ou misto.

Modelos mistos são utilizados para descrever dados de experimentos cuja estrutura de tratamentos envolve alguns fatores fixos e alguns aleatórios, ou seja, modelos lineares que contêm efeitos fixos e aleatórios, independentemente da média e do erro. Assim, esses modelos envolvem duas partes: uma parte descrevendo os efeitos aleatórios e a outra descrevendo os efeitos fixos. Conseqüentemente, a análise de um modelo misto consiste de duas partes: uma análise para a parte aleatória e outra para a parte fixa (PERRI e IEMMA, 1999).

Segundo Duarte e Vencovsky (2001), na área de genética e melhoramento, os genótipos avaliados devem ser considerados por definição, aleatórios, enquanto os demais fatores podem ser considerados fixos. Neste caso, a análise de experimentos por meio de um modelo misto é exigida. No entanto, estes mesmos autores relatam que tem sido comum o uso de análise baseada em modelo fixo para a estimação de médias de genótipos, mesmo quando estes genótipos foram obtidos por amostragem numa população. Nesse caso, a metodologia de modelos mistos é

geralmente utilizada, com o rigor da suposição, apenas para a estimação de componentes de variância e para a construção dos testes F apropriados à análise de variância.

Nos modelos mistos, a análise da parte aleatória consiste na predição dos efeitos aleatórios, na presença de efeitos fixos e na estimação dos componentes de variância. A análise da parte fixa consiste na estimação e testes de hipóteses sobre funções estimáveis dos efeitos fixos. Em geral, tanto a predição dos efeitos aleatórios quanto a estimação dos efeitos fixos dependem da estimação dos componentes de variância.

A metodologia de modelos mistos é uma ferramenta flexível para estimação de parâmetros genéticos e predição de variáveis aleatórias (valores genéticos) e apresenta várias vantagens (RESENDE, 2002). Dentre as vantagens apresentadas, podem-se destacar: pode ser aplicada a dados desbalanceados; permite utilizar simultaneamente um grande número de informações provenientes de diferentes gerações, locais e idades, gerando estimativas e predições mais precisas; não exige dados obtidos sob estruturas rígidas de experimentação, podendo ser aplicada a dados obtidos normalmente nos programas de melhoramento, os quais não precisam estar associados a delineamentos, bastando que se tenha informação sobre a genealogia dos indivíduos; permite a estimação de efeitos de dominância e epistáticos, além dos aditivos, pois utiliza maior número de relações de parentesco; e permite comparar indivíduos analisados ao longo do tempo e espaço. Além disso, permite prever o desempenho de híbridos não realizados, desde que informações de relacionamento genético entre as linhagens estejam disponíveis (BERNARDO, 1994,1995,1996; ANDRÉ, 1999; IEMMA, 2003).

Alguns métodos têm sido utilizados com a finalidade de prever o valor genotípico em plantas pelo uso de diferentes funções dos dados. Estes métodos diferem principalmente pelas pressuposições assumidas e pelas considerações realizadas sobre a natureza dos efeitos dos valores genotípicos (WHITE e HODGE, 1989). Os procedimentos mais usuais têm sido o *BLUE – Best Linear Unbiased Estimator* (melhor estimador linear não tendencioso), o *BLP – Best Linear Predictor* (melhor preditor linear) e, mais

recentemente, o *BLUP* – *Best Linear Unbiased Predictor* (melhor preditor linear não tendencioso) (NUNES, 2006).

O procedimento *BLUP* foi desenvolvido por Henderson, em 1949, no contexto de modelos mistos, com a finalidade de prever valores genéticos de animais para a produção de leite, sob condições de desbalanceamento nos dados (MRODE, 2005; WHITE e HODGE, 1989), porém, somente foi apresentado formalmente por Henderson em 1973 (RESENDE, 2002). Uma síntese histórica da evolução do procedimento *BLUP* é apresentada em Resende (2002).

Duarte e Vencovsky (2001) propuseram uma reflexão teórica sobre o processo de estimação/predição de médias de tratamentos no melhoramento vegetal. Neste sentido, compararam as análises baseadas no modelo linear fixo (análise intrablocos) e no modelo linear misto com genótipos aleatórios (análise recuperando informação intertratamentos), buscando identificar os fatores que podem determinar diferentes classificações genóticas. A análise teórica permitiu constatar que a abordagem de modelo misto (com tratamentos aleatórios), comparativamente às análises tradicionais (médias marginais e análise intrablocos), em geral, leva a uma maior homogeneidade das médias de tratamentos e seleção de diferentes tratamentos genéticos, quando a variância genotípica for baixa em relação à variância do erro e os ensaios forem não ortogonais e desbalanceados. Ademais, se os tratamentos forem oriundos de várias populações, a predição BLUP pode determinar diferente classificação das médias de tratamentos em relação à análise intrablocos, mesmo sob ortogonalidade e balanceamento.

2.2. Estimação de componentes de variância

Componentes de variância são definidos como variâncias associadas aos efeitos aleatórios de um modelo estatístico (BARBIN, 1993), sendo o seu conhecimento de crucial importância na área de genética e melhoramento, uma vez que as estratégias de seleção a serem utilizadas dependem de informações que podem ser obtidas a partir desses componentes.

Estimativas de variância e covariância são fundamentais para a teoria e aplicação do melhoramento genético, destacando-se os seguintes usos: cálculos de índices de seleção, planejamento de programas de melhoramento, interpretação de mecanismos genéticos de características quantitativas e estimação de coeficientes de herdabilidade e de correlações genética, fenotípica e ambiental (FREITAS, 1991).

De acordo com Fisher (1918), uma população é caracterizada pelas suas freqüências de genes, de genótipos e pelos seus valores genotípicos. Por outro lado, a teoria da genética quantitativa baseia-se em conceitos, tais como efeitos médios de uma substituição gênica, valores genéticos, variância aditiva e não aditiva, efeitos epistáticos, entre outros. Estes dependem de freqüências gênicas desconhecidas e valores genotípicos de características quantitativas que somente foram aplicados ao melhoramento após a estimação de componentes da variância genética, obtidas da similaridade entre parentes.

O progresso genético é dependente da fração da variância total que é de natureza genética. Adicionalmente, o conhecimento da magnitude relativa da parte aditiva e não aditiva da variância genética é fundamental no planejamento de programas de melhoramento. De acordo com Freeman (1979) e Henderson (1985a), o melhoramento genético de muitas características em animais é feito com base na variância aditiva, por meio de modelos estritamente aditivos. Em muitas características, a dominância é importante, tanto no melhoramento animal, quanto vegetal. Como exemplo, pode-se citar o melhoramento da produção em milho, em qual se busca a heterose por meio do cruzamento entre linhas contrastantes. No caso do melhoramento animal, podem-se citar os casos de irmãos completos em populações de suínos, em que a heterose tem sido demonstrada em várias características individuais.

O conhecimento prévio dos componentes de variância e covariância é necessário na predição dos valores genéticos quando se faz uso de métodos de predição como o BLUP. Porém, estes componentes não são geralmente conhecidos e podem ser estimados por vários métodos.

O procedimento usual para a obtenção dessas estimativas, quando se utilizam dados balanceados, é a análise de variância. Entretanto, no

melhoramento genético, muitas vezes, são usados dados desbalanceados, dentro dos efeitos fixos e aleatórios e contrastes não ortogonais. Neste caso, várias são as formas de obter as somas de quadrados dos efeitos e, conseqüentemente, a solução dos estimadores de variâncias.

Os estimadores ANOVA e seus derivados (Métodos I, II e III de Henderson) (HENDERSON, 1953) foram amplamente usados por causa das propriedades ótimas de seus estimadores, como a não-tendenciosidade e a variância mínima, dentre outras. Por outro lado, há possibilidades de obtenção de estimativas negativas e, com dados não balanceados, somente a não-tendenciosidade dos estimadores é reconhecida. Esta metodologia, apesar de ser relativamente fácil do ponto de vista computacional, tem sérios inconvenientes: não controla vícios das estimativas decorrentes de seleção; não permite a utilização do parentesco entre os indivíduos; e não é adequada para características múltiplas (FREITAS, 1991). Como conseqüência destas deficiências, o melhorista é obrigado a recorrer a técnicas de análise mais eficientes, a fim de obter dos dados as informações com mais rigor e precisão.

Ultimamente, devido ao aumento real de memória nos microcomputadores, grande ênfase tem sido dada aos métodos mais sofisticados de estimação de componentes de variância e covariância, tais como o método da máxima verossimilhança – ML (HARTLEY e RAO, 1967) e o método da máxima verossimilhança restrita – REML (PATERSON e THOMPSON, 1971). Esses métodos, apesar de extremamente trabalhosos, são superiores aos de HENDERSON, pois propiciam estimadores que são não-tendenciosos, assintoticamente eficientes e de translação constante. Além destas propriedades, estes métodos são adequados para características múltiplas, permitem a inclusão dos parentes dos indivíduos envolvidos, aplicam-se a qualquer matriz de incidência e controlam ou minimizam os vícios das estimativas devido à seleção (HENDERSON, 1984, 1985, 1986).

O método da Máxima Verossimilhança (ML), derivado por Hartley e Rao (1967), citados por Thompson (1979), utiliza-se de formas quadráticas calculadas, levando em consideração os próprios valores dos parâmetros. Portanto, as esperanças dessas formas quadráticas não são lineares nos

parâmetros e devem ser solucionadas iterativamente, o que dificulta a obtenção das estimativas. Por outro lado, o uso dessas formas quadráticas mais complexas é efetivo, já que, segundo Rothschild *et al.* (1979), elas fornecem estimativas não-viesadas dos componentes de variância na presença de seleção, o que não ocorre com o Método I de Henderson.

Para obtenção dos estimadores por ML, é necessário o conhecimento da distribuição da variável em estudo. De acordo com Anderson (1984), este método é iterativo, fornece estimativas não-negativas dos componentes de variância e elimina o “viés” atribuído às mudanças nas frequências gênicas resultantes da seleção, porém não considera a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo. A estimação por meio do método ML requer a inversão da matriz da parte de efeitos aleatórios das equações de modelos mistos, e isto pode ser computacionalmente difícil ou aproximadamente impossível para esquemas complexos que envolvem vários fatores aleatórios com números grandes de níveis, e isso, segundo Blasco (2001), o ML não foi utilizado para determinação do valor genético até recentemente. Outra limitação do ML é de que a normalidade é requerida para que os estimadores tenham a propriedade de máxima verossimilhança (KENNEDY, 1991).

Patterson e Thompson (1971) desenvolveram o método atualmente utilizado, denominado método da Máxima Verossimilhança Restrita (REML). A diferença deste para o método ML é que, ao invés de utilizar toda a função de verossimilhança, ele se utiliza apenas dos termos da verossimilhança relativos a contrastes ortogonais da parte aleatória das observações, pois os contrastes entre os efeitos fixos não fornecem nenhuma informação adicional sobre o resíduo ou sobre os efeitos aleatórios. O método REML tem a vantagem de dar o mesmo resultado da ANOVA em dados balanceados (ANOVA é não viesada, com variância mínima, para tal situação).

De acordo com Meyer (1986), o método da máxima verossimilhança restrita (REML) é o mais recomendado para dados desbalanceados. A mesma autora considera o método REML mais adequado para estimação de parâmetros genéticos, pois além de considerar a perda de graus de liberdade resultante da estimação dos efeitos fixos, as estimativas caem

sempre dentro do espaço paramétrico. Esse método evita erros de pequenas amostras associadas aos efeitos fixos e as dados provenientes de populações submetidas à seleção. Os métodos REML e ML têm a notável propriedade na estimação de parâmetros da população base, livre de viés, devido à seleção (HENDERSON, 1986).

Existem alguns autores que questionam a superioridade do método REML, mas Gianola e Fernando (1986), utilizando-se de argumentos Bayesianos, apresentam razões consistentes para que o método REML seja considerado superior ao ML. Segundo esses autores, os chamados efeitos fixos são parâmetros de distúrbio quando os interesses são os componentes de variância dos efeitos aleatórios, e o que se faz, quando se excluem os contrastes dos efeitos fixos e se utiliza apenas a verossimilhança dos contrastes entre as partes aleatórias das observações, é eliminar os parâmetros de distúrbio por meio de integração. Isto equivale a fazer inferências sobre a distribuição marginal dos componentes de variância e covariância, o que é coerente, visto que esses componentes são obtidos numa etapa anterior e, portanto, devem ser válidos para quaisquer que sejam os valores dos efeitos fixos a serem estimados.

Apesar de a pressuposição de normalidade dos dados ser crucial para o procedimento de estimação por ML e REML, a literatura não é informativa sobre como o não atendimento desta exigência influencia as estimativas de parâmetros genéticos no melhoramento (FREITAS, 2000). Na análise de variância, entretanto, há evidências de que os valores da curtose e, em menor extensão, da assimetria dos dados (SCHEFFÉ, 1959) são os indicadores mais importantes de como os desvios de normalidade dos dados influenciam as inferências obtidas. Segundo Cochran e Cox (1978), a não-normalidade dos dados pode influir negativamente na estimação dos efeitos fixos, no uso dos testes t e F e na heterogeneidade da variância do erro.

Assumir esta hipótese na estrutura de dispersão dos dados, quando a mesma não é atendida, implica prejuízos na resposta à seleção (GIANOLA *et al.*, 1992).

2.3. BLUP no melhoramento genético vegetal

Muitos trabalhos envolvendo BLUP no melhoramento vegetal têm sido realizados nos últimos anos, envolvendo principalmente análises dialélicas e predição de híbridos não realizados e também a estimação de componentes de variância e identificação de indivíduos geneticamente superiores. PIEPHO *et al.* (2008) analisaram a evolução recente da metodologia BLUP aplicada ao melhoramento de plantas. Entre as aplicações, apresentam modelos que levam em consideração informações genealógicas, explorando a correlação genética entre parentes, e também modelos que não consideram a matriz de parentesco. Esses autores demonstraram que o BLUP é mais preciso quando comparado aos outros procedimentos de predição e estimação e destacam a potencialidade da metodologia aplicada ao melhoramento vegetal.

Como no melhoramento genético é comum o interesse em mais de uma característica, Bauer e Leon (2008) desenvolveram um trabalho de simulação que analisasse o mérito da seleção de linhas superiores baseada no BLUP multicaracterístico. Foram simuladas características com diferentes valores de herdabilidade e correlação. Maiores ganhos devidos à seleção foram observados para BLUP multi-característico em relação às análises que consideraram apenas uma característica, principalmente quando a correlação entre as duas características era negativa.

Bernardo (1994), usando a metodologia de modelos mistos, com estimativas de parentesco obtidas por marcadores moleculares do tipo RFLP, para prever o desempenho de alguns híbridos não realizados, obteve estimativas de correlação entre o valor predito e o desempenho do cruzamento em campo, para produção de grãos, variando de 0,65 a 0,80. A eficiência na predição de cruzamentos não realizados aumentou à medida que se empregou maior número de cruzamentos como preditores. Isso mostra a variabilidade da utilização dessa metodologia como forma de identificação de cruzamentos promissores antes de sua realização.

Bernardo (1995) avaliou alguns modelos genéticos para realizar predições, considerando graus elevados de desbalanceamento do dialelo. Os valores de correlação entre o desempenho observado e predito para a

produção de grãos variaram entre 0,58 e 0,75, quando foi utilizado um grupo de 67 híbridos. Os valores de correlação foram considerados altos mesmo quando em situações em que o número de preditores era menor que 20% do número de híbridos com desempenho a ser predito. Estes resultados sugerem que o BLUP pode ser eficiente na obtenção do valor genético de um número elevado de híbridos, avaliando-se um número bem menor de cruzamentos que o utilizado geralmente nos programas de melhoramento.

Bernardo (1996), utilizando um conjunto de 4099 híbridos de 16 combinações entre nove grupos heteróticos de milho, avaliados em até 15 locais, encontrou estimativas de correlação entre valor predito e o desempenho fenotípico, para produção de grãos, variando de 0,49 a 0,70, com uma tendência de estimativas mais elevadas para os grupos com maior número de híbridos avaliados. Segundo o autor, estes valores podem ser considerados moderadamente elevados e sugerem que o BLUP seja empregado de forma eficiente para identificação de cruzamentos não realizados, com desempenho superior. Entretanto, sua utilização em larga escala, como método preditivo de cruzamentos não realizados, necessita ser mais bem pesquisada, principalmente no tocante à quantidade e ao tipo de cruzamentos a serem utilizados como preditores, bem como aos modelos genéticos mais adequados.

Charcosset *et al.* (1998) analisaram diversos modelos para predição de híbridos F1 de milho, encontrando como resultado elevada eficiência de diversos modelos. Os autores afirmam que, em situações práticas, a escolha do modelo depende do parentesco entre as linhagens.

Em estudo de resistência do milho a patógenos de raiz, Moreno-Gonzales *et al.* (2004) compararam quatro modelos genéticos e dois procedimentos estatísticos: o modelo misto para a melhor predição linear não viciada (BLUP) e o modelo linear geral (GLM) na predição dos efeitos de capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC). Para CGC, houve boa concordância entre os dois modelos estudados. No entanto, a seleção com base apenas nas linhagens não conferiu resistência aos híbridos.

Kalil Filho *et al.* (2000), utilizando a metodologia BLUP/REML, estimaram parâmetros genéticos e obtiveram as predições dos valores

genéticos aditivos e genotípicos de clones amazônicos de seringueira em relação às características de qualidade da borracha e do látex. O trabalho foi realizado a partir de genealogia envolvendo 15 clones e 250 rametes. As estimativas de herdabilidade no sentido amplo foram de 0,4152 e 0,6357, com relação a teores de borracha seca e plasticidade após 30 minutos, respectivamente, o que torna possível obter altos ganhos genéticos via seleção clonal.

Com o objetivo de aplicar o método REML/BLUP em programas de melhoramento genético do cafeeiro, utilizando-o na estimação de parâmetros genéticos e predição de valores genotípicos, Resende *et al.* (2001) demonstraram ser o método adequado à estimação de parâmetros genéticos e predição de valores genotípicos, podendo ser empregado rotineiramente no melhoramento do cafeeiro.

Trabalhando com melhoramento de pupunha para produção de palmito, Farias Neto e Resende (2001) estudaram a variabilidade genética e estimaram o valor genético individual como critério de seleção, usando o procedimento BLUP/REML. Adotaram-se duas estratégias de seleção: simulando programa de melhoramento a curto prazo (CP- seleção das 9 famílias com 31 indivíduos de maior valor genético) e a longo prazo (LP- seleção das 15 famílias com 53 indivíduos). Os ganhos genéticos preditos em relação à média da população, para peso líquido de palmito, foram de 7,18% na situação de LP e de 8,40% para CP.

Com o objetivo de selecionar os melhores clones de eucalipto aos sete anos de idade, Garcia e Nogueira (2005) utilizaram-se da metodologia de modelos mistos para a obtenção da Melhor Predição Linear não Viesada (BLUP) dos efeitos genotípicos e o processo da Máxima Verossimilhança Restrita (REML) para a estimação dos componentes de variância e dos parâmetros genotípicos. O processamento foi realizado utilizando o programa SELEGEN, cujo vetor de soluções apresenta os efeitos genotípicos preditos e ganhos de seleção para cada clone. A seleção dos clones com desempenho superior a 80% resultou numa nova média para volume, de 0,519 m³, correspondendo a um aumento de 22,8% em relação à média das testemunhas, que foi de 0,415 m³, e ganho de 36,8% em relação à média do experimento, que foi de 0,380 m³.

Na avaliação genética de progênies de meio-irmãos de *Eucalyptus urophylla* S. T. Blake, Rocha *et al.* (2006) utilizaram os procedimentos REML/BLUP e o método dos quadrados mínimos. Nos cinco testes avaliados aos 58 meses de idade, para a característica diâmetro à altura do peito (DAP), foram encontrados valores dentro dos padrões da espécie, além de exibir variabilidade genética significativa, pelo teste F, a 1% de probabilidade. Nos dois procedimentos, máxima verossimilhança restrita (REML) e método dos momentos, os valores dos parâmetros genéticos foram semelhantes, exceto entre famílias, em que o procedimento REML proporcionou valores mais elevados, com acurácia superior a 70% em todas as populações, mostrando-se uma ferramenta apropriada para esse fim.

Naoe *et al.* (2006), com objetivo de verificar a influência do parentesco na estimação de componentes de variância e obter parâmetros genéticos para conteúdo protéico nas sementes de soja, em programa de retrocruzamento, utilizaram o método da máxima verossimilhança restrita. O modelo utilizado considerou fixas as diferentes épocas de plantio e aleatórios os efeitos genéticos. Esses autores perceberam que, incorporando informação de geração descendente ou ascendente, ocorreram mudanças nas estimativas dos componentes de variância e na herdabilidade. Entretanto, sua magnitude variou dependendo das particularidades de cada população estudada.

Farias Neto *et al.* (2007) fizeram a avaliação genotípica de progênies e indivíduos de açaí para vários caracteres. Foram apresentadas também estimativas de parâmetros genéticos e ambientais que permitiram inferir sobre o controle genético e nível de variabilidade genética presente no material avaliado, assumindo um sistema reprodutivo misto, com a ocorrência de 12,9% de autofecundação. Para estimação dos componentes de variância e predição dos valores genéticos foi empregada a metodologia REML/BLUP. Verificou-se que, de um modo geral, todos os caracteres apresentaram variabilidade genética expressiva, evidenciando o grande potencial para a seleção na própria população experimental. Dentre as estratégias de seleção estudadas, a seleção individual apresentou maior estimativa de ganho genético relativamente ao estimado para a seleção de progênies. Usando técnicas de análise multivariada, foram obtidos seis

grupos divergentes, que podem orientar cruzamentos visando à busca de descendentes heteróticos.

Segundo Coimbra *et al.* (2005), na área de genética e melhoramento é comum conduzir experimentos constituídos de fatores fixos (método de semeadura) e aleatórios (populações). Sendo assim, a análise estatística mais apropriada deveria ser por meio de modelos lineares mistos. Neste contexto, os autores propuseram um trabalho com os objetivos de estimar os componentes de variância para os efeitos do fator aleatório e interação população x método de semeadura, realizar a seleção entre populações e entre populações dentro de cada método de semeadura e comparar o efeito do fator fixo. Este trabalho revelou que, tanto o efeito dos componentes da variância das populações quanto o efeito da interação simples população x métodos de semeadura revelaram uma baixa contribuição para a variância total. Desse modo, pode-se concluir que as populações segregantes apresentam uma estreita variabilidade genotípica, sendo o procedimento Reml/Blup indicado para a estimação/predição no melhoramento de plantas de aveia, cujos experimentos, em geral, geram dados não balanceados.

Recentemente, diferentes aplicações têm sido apresentadas à metodologia de modelos mistos no melhoramento genético, como é caso do trabalho de Campos *et al.* (2006), que utilizaram a abordagem de modelos mistos para avaliar o progresso genético de híbridos de milho em relação à tolerância à seca. Outro trabalho foi o de Smith *et al.* (2006), que propuseram a avaliação de experimentos desbalanceados com substancial fonte de correlação e variação não genética, por meio de modelos mistos. O estudo de simulação mostrou que o uso de análises por modelos mistos apresentou vantagens em termos de ganhos devidos à seleção.

Segundo Bauer *et al.* (2006), a utilização de informação de parentesco na predição de valores genéticos em plantas autógamas tem algumas desvantagens, principalmente quando a informação de pedigree não está completa. Por esse motivo, os autores propuseram o cálculo da matriz de parentesco com base em informações de marcadores moleculares utilizando dados de cevada. Os resultados obtidos indicaram que os dados de marcadores moleculares podem ser utilizados na obtenção da matriz de parentesco.

Como no melhoramento genético é comum o interesse em mais de uma característica, Bauer e Leon (2008) desenvolveram um trabalho de simulação que analisasse o mérito da seleção de linhas superiores baseada no BLUP multicausal. Foram simuladas características com diferentes valores de herdabilidade e correlação. Maiores ganhos devido à seleção foram observados para BLUP multicausal em relação às análises que consideraram apenas uma característica, principalmente quando a correlação entre as duas características era negativa.

Em um contexto diferente dos trabalhos citados anteriormente, Gonçalves *et al.* (2007) desenvolveram uma pesquisa com o objetivo de demonstrar a relevância da análise espacial com base em modelos mistos, no melhoramento de videira. Ficou demonstrado que todos os modelos de análise espacial (gaussiano, esférico, potência e potência anisotrópico) foram superiores ao de blocos ao acaso, uma vez que apresentavam menores valores dos critérios de informação de Akaiake e Bayesiano. As estimativas de variância genotípica foram equivalentes nos diferentes modelos. Apesar de as correlações entre os valores genotípicos preditos pelos distintos modelos serem altas (0,946 a 0,979), houve alterações de rank, e as predições foram mais precisas com o modelo de análise espacial. Assumindo 15% de clones selecionados, as diferenças entre os grupos de selecionados com análise espacial e pela análise de modelo em blocos variaram de 10% a 29,7%. Os ganhos preditos com a análise espacial foram um pouco superiores àqueles preditos com ajuste do modelo de blocos e mais precisos. Entre os modelos de análise espacial, destacou-se o potência anisotrópica. Sebolai *et al.* (2005), por meio de simulação de dados, realizaram análise espacial utilizando equações de modelos mistos em testes de cultivares precoces. Ficou evidenciado no trabalho que o uso de BLUP fornece resultados favoráveis, permitindo a seleção de genótipos precoces verdadeiramente superiores.

3. MATERIAL E MÉTODOS

3.1. Metodologia de modelos mistos

Em programa de melhoramento populacional recorrente com famílias de meios-irmãos, a superioridade genotípica de uma progênie é determinada pela superioridade genética (número de genes favoráveis) do seu genitor comum. Em população em equilíbrio de Hardy-Weinberg, a média genotípica de uma família de meios-irmãos é (VIANA, 2004), admitindo ausência de epistasia,

$$\bar{G}_{HSF} = M + \frac{1}{2}A$$

sendo M a média populacional e A o valor genético aditivo do pai comum, um indicador quase perfeito de sua superioridade quanto ao número de genes favoráveis (LYNCH e WALSH, 1988). A correlação entre o valor genético aditivo do genitor comum e a freqüência média dos genes favoráveis é, assumindo equilíbrio da fase gamética,

$$\rho_{A, \bar{p}} = \frac{(1/2k) \sum_{i=1}^k p_i q_i \alpha_i}{\sqrt{\sigma_A^2 \left[(1/8k^2) \sum_{i=1}^k p_i q_i \right]}}$$

em que k é o número de genes, p é a freqüência do gene favorável, α_i é o efeito de substituição gênica e $\sigma_A^2 = 2 \sum_{i=1}^k p_i q_i \alpha_i^2$ é a variância genética aditiva.

Em relação a um gene, $\rho_{A, \bar{p}} = 1$, com dominância completa, com dominância parcial e com ausência de dominância. No caso de sobredominância, $\rho_{A, \bar{p}} = 1$ ou -1 , dependendo da freqüência do dominante na população.

Portanto, pode-se proceder à seleção das progênies superiores a partir da predição dos valores genéticos aditivos dos pais comuns, na presença de um ou mais efeitos fixos, empregando-se a metodologia proposta por Henderson (1974). O procedimento usual dos melhoristas é

selecionar com base no BLUE da média genotípica, estimada também na presença de um ou mais efeitos fixos, empregando o processo de estimação de quadrados mínimos (ordinário, ponderado ou generalizado, dependendo das pressuposições acerca da matriz de covariâncias residuais). Em geral, admitindo apenas um efeito aleatório [o efeito do pai comum, ou seja, $(1/2)A$], o valor fenotípico de uma família é

$$P_{HSF} = M + \frac{1}{2}A + \sum_i (\text{efeito fixo})_i + \text{erro}$$

Em termos matriciais, tem-se

$$y = X\beta + Zu + e$$

em que y é o vetor de valores fenotípicos, X é a matriz de incidência dos efeitos fixos, β é o vetor dos níveis de cada fator fixo (incluindo a média populacional), Z é a matriz de incidência dos efeitos aleatórios, u é o vetor dos valores genéticos aditivos dos pais comuns, divididos por 2, e e é o vetor de resíduos.

O critério de obtenção do melhor preditor linear imparcial de u (BLUP(u)) é a maximização da função densidade probabilidade conjunta de y e u (HENDERSON, 1950), obtendo, sob normalidade, as equações de modelo misto (MME) de Henderson. A função de verossimilhança é

$$f(y, u) = f(y|u).f(u) = \frac{1}{(2\pi)^{(n/2)} |R|^{(1/2)}} e^{-\frac{1}{2}[(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu)]} \cdot \frac{1}{(2\pi)^{(n/2)} |G|^{(1/2)}} e^{-\frac{1}{2}[(u - \phi)'G^{-1}(u - \phi)]}$$

sendo R a matriz de covariâncias residuais e G a matriz de covariâncias de u ($e \sim N(\phi, R)$ e $u \sim N(\phi, G)$, pois $E(A) = 0$). A função suporte é

$$\ln f(y, u) = -2\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |R| - \frac{1}{2} \ln |G| - \frac{1}{2} \cdot$$

$$\left(y'R^{-1}y - 2y'R^{-1}X\beta - 2y'R^{-1}Zu + \beta'X'R^{-1}X\beta + 2\beta'X'R^{-1}Zu + u'Z'R^{-1}X\beta + u'G^{-1}u \right)$$

Derivando em relação a β e a u e tornando as derivadas identicamente nulas, resulta que

$$\begin{bmatrix} -2 \frac{\partial \beta'}{\partial \beta} X'R^{-1}y + 2 \frac{\partial \beta'}{\partial \beta} X'R^{-1}X\beta^0 + 2 \frac{\partial \beta'}{\partial \beta} X'R^{-1}Z\tilde{u} \\ -2 \frac{\partial u'}{\partial u} Z'R^{-1}y + 2 \frac{\partial u'}{\partial u} Z'R^{-1}X\beta^0 + 2 \frac{\partial u'}{\partial u} Z'R^{-1}Z\tilde{u} + 2 \frac{\partial u'}{\partial u} G^{-1}\tilde{u} \end{bmatrix} = \begin{bmatrix} \phi \\ \phi \end{bmatrix}$$

Ou ainda

$$\begin{bmatrix} X'R^{-1}X\beta^0 & X'R^{-1}Z\tilde{u} \\ Z'R^{-1}X\beta^0 & (Z'R^{-1}Z + G^{-1})\tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

finalmente,

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

A matriz de covariâncias de u é

$$G = (1/4)\sigma_A^2 \begin{bmatrix} 1 & 2r_{12} & \cdots & 2r_{1t} \\ 2r_{12} & 1 & \cdots & 2r_{2t} \\ \cdots & \cdots & \ddots & \cdots \\ 2r_{1t} & 2r_{2t} & \cdots & 1 \end{bmatrix} = (1/4)\sigma_A^2 A$$

pois $\text{Cov}(u_i, u_j) = \text{Cov}(\frac{1}{2}A_i, \frac{1}{2}A_j) = \frac{1}{4} \cdot 2r_{ij}\sigma_A^2$, sendo r_{ij} o coeficiente de

parentesco entre os genitores das progênieis i e j (LYNCH e WALSH, 1988).

A matriz A é denominada matriz de parentesco genético aditivo. Assumindo

$R = I\sigma^2$, resulta que:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A \end{bmatrix} \begin{bmatrix} \beta^0 \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

em que $\lambda = \sigma^2 / (1/4)\sigma_A^2$.

A metodologia REML/BLUP consiste em, de maneira iterativa, obter a solução para o vetor de efeitos fixos, prever os efeitos aleatórios e estimar os componentes de variância (HENDERSON, 1984; MRODE, 2005). Ao final do processo iterativo, com a convergência dos componentes de variância,

$$\begin{bmatrix} \beta^\circ \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \hat{\lambda}A^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} X'y \\ Z'y \end{bmatrix} = C \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

em que $\hat{\lambda} = \tilde{\sigma}^2 / (1/4)\tilde{\sigma}_A^2$, sendo $\tilde{\sigma}^2$ e $(1/4)\tilde{\sigma}_A^2$ os estimadores de máxima verossimilhança da variância residual e da variância genotípica entre progênieis.

O estimador de herdabilidade em nível de média de famílias é

$$\hat{h}^2 = \frac{(1/4)\tilde{\sigma}_A^2}{\tilde{\sigma}^2/k + (1/4)\tilde{\sigma}_A^2}$$

sendo o valor de k dependente do delineamento experimental.

Para testar q combinações lineares estimáveis ($H_0 : L\theta = \phi$, $\theta' = [\beta' \ u']$), envolvendo efeitos fixos e, ou, aleatórios, a estatística é (LITTELL *et al.*, 2006)

$$F = \frac{\hat{\theta}'L'(L'CL)^{-1}L\hat{\theta}}{q}$$

associada a $q =$ posto de L e v graus de liberdade.

As principais propriedades do BLUP de u são (HENDERSON, 1984; SEARLE *et al.*, 1992)

- minimiza a variância do erro de predição ('BEST');
- é uma função linear de y ('LINEAR');
- é estimador imparcial de u ('UNBIASED');
- em caso de distribuição normal multivariada, o BLUP de u é BLUE, MLE ('maximum likelihood estimator') e estimador bayesiano de u ; e
- em caso de distribuição normal multivariada, a função linear que maximiza a probabilidade de selecionar o melhor entre u_i e u_j é BLUP(u_i) – BLUP(u_j).

3.1.1. Análises computacionais utilizando modelos mistos

Os procedimentos computacionais utilizados nas análises de modelos mistos foram elaborados utilizando o software SAS[®], por meio dos procedimentos *proc inbreed*, *proc iml* e *proc mixed*. Embora apresentadas de forma separada para facilitar a compreensão do leitor, essas rotinas devem ser processadas em um único comando seguindo a seqüência na qual estão apresentadas.

Na análise de modelos mistos, é necessário que se calcule a matriz de parentesco entre os indivíduos avaliados. Nesse caso, a matriz de parentesco foi calculada pelo *proc inbreed* do software SAS[®]. Para que o software calcule essa matriz, é de fundamental importância informar corretamente o pedigree.

A forma correta de montar um pedigree para que o software processe os dados é informar, em três colunas, o indivíduo, o pai e a mãe. Os primeiros indivíduos a serem inseridos no pedigree são aqueles que não têm informação sobre sua ancestralidade, preenchendo com ponto (.) os locais dos pais e, ou, das mães que não são conhecidos.

No caso de seleção recorrente entre famílias de meios-irmãos, não existe parentesco entre as famílias do primeiro ciclo. Portanto, tanto a coluna referente aos pais, quanto a referente às mães serão preenchidas por pontos, desse modo, a matriz de parentesco calculada será uma identidade. Os demais indivíduos, por se tratar de meios-irmãos, apresentarão informações apenas em uma das colunas: na coluna das mães, se meios-irmãos maternos, ou na dos pais, se forem meios-irmãos paternos. A outra coluna deve ser também preenchida com pontos.

Depois de inserir os indivíduos do primeiro ciclo de seleção no pedigree, as próximas informações a serem acrescentadas são das plantas do lote de recombinação desse ciclo. Essas plantas são filhas das famílias testadas no ciclo anterior e são mães das famílias do próximo ciclo. Nesse caso, coloca-se o número referente à planta na primeira coluna, a segunda coluna será preenchida por ponto e na terceira coluna será colocado o número da família do ciclo anterior que deu origem à planta do lote de recombinação. Em seguida, podem-se alocar no pedigree as famílias

testadas no segundo ciclo de seleção. Na primeira coluna, coloca-se o número da família testada, a segunda preenche-se com ponto, já que não há informação sobre os pais, e na terceira coluna informa-se o número referente à planta-mãe, que foi informada anteriormente no lote de recombinação. Para os ciclos seguintes, deve-se realizar o mesmo procedimento.

Um exemplo de pedigree utilizado na avaliação de três ciclos de seleção recorrente entre famílias de meios-irmãos de milho-pipoca pode ser visto no Anexo 1.

Neste exemplo, os indivíduos de 1 a 196 apresentados na primeira coluna são as famílias do primeiro ciclo de seleção, portanto, a segunda coluna, que é referente aos pais, e a terceira, referente às mães não possuem informação e estão preenchidas com ponto. Os indivíduos de 197 a 392 e de 589 a 784 são as plantas do lote de recombinação do primeiro e segundo ciclos de seleção, respectivamente. Como são meios-irmãos maternos, a coluna dos pais está preenchida por ponto e a coluna das mães preenchida com números referentes às famílias do ciclo anterior. Os indivíduos de 393 a 588 e de 785 a 980 são as famílias do segundo e terceiro ciclos de seleção, respectivamente. A coluna dos pais está preenchida com ponto e a das mães preenchida com o número das plantas do lote de recombinação do ciclo anterior.

Para obter a matriz de parentesco utilizando o *proc inbreed* do SAS deve-se seguir a rotina a seguir:

```
data teste;
input ind pai mae;
datalines;
(Arquivo contendo pedigree)
proc inbreed data=teste covar noprint outcov=saida0;
var ind pai mae;
run;
data matriz;set saida0(firstobs=*);keep COL**-COL***;
```

* neste local, deve ser colocado o número referente ao primeiro indivíduo do teste para o qual se deseja obter a matriz de parentesco;

** e *** nestes locais devem ser colocados os números do primeiro e do último indivíduo, respectivamente, do teste para o qual se deseja obter a matriz de parentesco.

Para ser lida pelo *proc mixed* do SAS®, a matriz de parentesco necessita estar em uma forma especial de apresentação. Essa matriz deve ser transformada em um arquivo no formato de matriz com as colunas *parm*, *row*, *col*, *value*, em que a coluna *parm* deve ser preenchida com o número 1, a coluna *row* corresponde à linha referente ao indivíduo *i* da matriz de parentesco, a coluna *col* corresponde à coluna referente ao indivíduo *j* da matriz de parentesco e a coluna *value* se refere ao coeficiente de parentesco entre os indivíduos *i* e *j*. Uma maneira de transformar essa matriz é utilizar o *proc iml* do SAS® por meio da seguinte rotina:

```
proc iml;
create saida1 var{B};
use matriz;
read all into A;
r=ncol(A)**2;
B=shape(A,r,1);
append var{B};
run;
data saida2;
do parm=1 to 1;
do row= # to # #;
do col=# to # #;output;
end;
end;
end;
run;
data A_parent;merge saida2 saida1;rename B=value;
run;
```

neste local deve ser colocado o número referente ao primeiro indivíduo do arquivo de dados referente ao teste para o qual se deseja obter a matriz de parentesco.

neste local deve ser colocado o número do último indivíduo do arquivo de dados referente ao teste para o qual se deseja obter a matriz de parentesco.

Os procedimentos para estimação dos componentes de variância, predição dos valores aleatórios e estimação dos efeitos fixos foram realizados no *proc mixed* do SAS[®]. O primeiro passo é informar o conjunto de dados a ser analisado de acordo com o modelo escolhido, como a seguir.

```
data teste1;
input (descrição da ordem dos dados no arquivo de entrada);
datalines;
(Arquivo de dados)
run;
```

Depois de realizada a entrada dos dados, pode-se proceder às análises do *proc mixed*, de acordo com a rotina apresentada abaixo:

```
proc mixed data=teste1 METHOD=REML COVTEST;
class ef1...efn ea1;
model v1 = ef1...efn / solution;
random ea1 / solution type=lin(1) ldata=A_parent;
parms (*) (**);
run;
estimate 'Media BLUP trat 1' intercept 1 | trat 1 ... 0
.
.
.
estimate 'Media BLUP trat n' intercept n | trat 0 ... n
;
run;
```

Em que

$ef_1...ef_n$ são os efeitos fixos do modelo (esse número de efeitos fixos pode variar),

ea_1 é o efeito aleatório do modelo,

A_{parent} é a matriz de parentesco gerada pelo *proc inbreed* e reestruturada pelo *proc iml*,

(*) e (**) são os valores iniciais para variância genotípica e residual respectivamente.

n é o número referente ao último tratamento testado

As saídas do *proc mixed* fornecem as estimativas dos componentes de variância genotípica e residual por REML, com seus respectivos desvios padrão e teste z para testar a hipótese de que os componentes de variância são iguais a zero, com sua respectiva probabilidade. É fornecido o teste z devido à propriedade de os estimadores de máxima verossimilhança seguirem distribuição normal. Um exemplo da saída do *proc mixed* pode ser observada no Anexo 2.

Quanto aos valores genéticos preditos, eles são fornecidos de duas maneiras. Primeiro se tem a saída dos valores da metade do valor genético do pai comum (por se tratar de famílias de meios-irmãos). Entretanto, esses valores podem ser tanto positivos quanto negativos. Para facilitar a interpretação, o programa fornece os valores preditos acrescentados à média. Isso torna a avaliação dos indivíduos mais simples, pois médias genotípicas das famílias passam a apresentar a mesma escala da característica avaliada. Essa saída com os valores acrescentados pela média é possível ao se utilizar a opção 'estimate' apresentada na rotina.

3.2. Método dos mínimos quadrados

O Método dos Mínimos Quadrados é uma técnica de otimização matemática que procura encontrar o melhor ajustamento para um conjunto de dados, tentando minimizar a soma dos quadrados das diferenças entre a curva ajustada e os dados (tais diferenças são chamadas resíduos). Um requisito implícito para o método dos mínimos quadrados é que os erros em cada medida sejam distribuídos aleatoriamente com função densidade gaussiana e que os resíduos sejam independentes. O Teorema de Gauss-Markov garante (embora indiretamente) que o estimador de mínimos quadrados (EMQ) é o estimador não-viesado de variância mínima linear na variável resposta.

Em um experimento, a resposta à seleção pode ser estimada usando procedimentos de quadrados mínimos como média das progênies dos pais

selecionados, sendo, neste caso, os efeitos ambientais estimados pelo uso de uma população controle não selecionada (BAKER *et al.*, 1991; PARNELL *et al.*, 1997; RAZOOK *et al.*, 1998). Em contraste com o estimador de modelo misto (considerando o modelo animal), o de quadrados mínimos não usa registros genealógicos para calcular a média da geração, ou seja, a análise de quadrados mínimos ignora a estrutura dos dados, usando somente os registros de uma geração em particular para computar a resposta e ponderar igualmente todos os indivíduos selecionados (WALSH, 1999).

A estimação dos componentes de variância se faz associando o quadrado médio à sua respectiva esperança matemática. O Quadro 1 apresenta o esquema da análise de variância para experimento em látice quadrado, cujo modelo matemático é,

$$Y_{ijk} = \mu + \tau_i + r_j + (b/r)_{jk} + e_{ijk}$$

em que,

Y_{ijk} é o valor fenotípico da família;

μ é a média geral;

τ_i é o efeito da progênie i (aleatório);

r_j é o efeito da repetição j (fixo);

$(b/r)_{jk}$ é o efeito do bloco k dentro da repetição j (fixo); e

e_{ijk} é o erro experimental.

Quadro 1 - Esquema da análise de variância de um experimento em látice quadrado, com as esperanças dos quadrados médios

F.V.	G.L.	Q.M.	E (Q.M.)
Repetição	$r-1$	QM Rep	$\sigma^2 + sv\phi_r$
Blocos/Rep.	$r(s-1)$	QMB	$\sigma^2 + s\phi_b$
Famílias	$v-1$	QMT	$\sigma^2 + [s/(s+1)]r\sigma_g^2$
Resíduo	$(s-1)(rs-s-1)$	QMR	σ^2

Os estimadores componentes de variância (VIANA e REGAZZI, 1999) são:

$$\hat{\sigma}^2 = \text{QMR}$$

$$\hat{\sigma}_g^2 = \left(\frac{s+1}{sr} \right) (\text{QMT} - \text{QMR})$$

em que,

σ^2 é o componente de variância residual;

σ_g^2 é o componente da variância genotípica;

r é o número de repetições; e

s é o número de blocos dentro de cada repetição.

De acordo com Falconer e Mackay (1996), a importância relativa do genótipo é dada pela razão entre a variância genotípica e a variância fenotípica. Sendo assim, a variância genotípica é a parte da variância fenotípica atribuída aos efeitos genéticos, já que ela resulta da ação conjunta do genótipo e do ambiente. A variância genotípica é composta pelas variâncias genéticas aditiva, devido à sua dominância, e epistática. A variância aditiva é o componente mais importante, uma vez que é a principal causa da semelhança entre parentes e, por conseguinte, a principal determinante das propriedades genéticas da população e da resposta da população à seleção (FALCONER e MACKAY, 1996).

A herdabilidade de um caráter quantitativo é uma das suas propriedades mais importantes. Além de expressar a proporção da variância total que é atribuída aos efeitos médios dos genes, que determinam o grau de semelhança entre parentes, ela possui função importante no estudo genético do caráter, com papel preditivo, expressando a confiança do valor fenotípico como um guia para o valor genético (FALCONER e MACKAY, 1996).

O estimador da herdabilidade em nível de média de famílias é:

$$\hat{h}^2 = \hat{\sigma}_g^2 / \hat{\sigma}_p^2 = (1/4) \hat{\sigma}_A^2 / \hat{\sigma}_p^2,$$

sendo $\hat{\sigma}_p^2$ o estimador da variância fenotípica, dado por:

$$\hat{\sigma}_p^2 = \hat{\sigma}_g^2 + \frac{1}{k} \hat{\sigma}^2$$

em que k é dependente do delineamento experimental.

Com as estimativas das médias genóticas das famílias, juntamente com o quadrado médio do resíduo, é possível construir intervalos de confiança para essas estimativas por meio de um valor crítico da estatística t associada ao número de graus de liberdade do resíduo a um nível crítico de probabilidade $\alpha/2$ igual a 2,5%. O intervalo é construído seguindo a expressão:

$$\bar{G}_{FMI} \pm t_{gl, \alpha/2} \sqrt{\hat{\sigma}^2/k}$$

3.3. Ganhos de seleção

A resposta à seleção direta em um caráter é uma função do diferencial de seleção (DS), da herdabilidade (h^2) e do controle parental adotado pelo esquema seletivo (p) (EBERHART, 1970). A função de predição de ganho é, genericamente,

$$\Delta G = p \cdot DS \cdot h^2 = p \cdot i \cdot \hat{\sigma}_{US}^2 / \hat{\sigma}_P,$$

sendo i a intensidade de seleção e $\hat{\sigma}_{US}^2$ a variância genotípica entre as unidades de seleção.

O diferencial de seleção é definido por:

$$DS = \hat{\mu}_{ps} - \hat{\mu}_p,$$

sendo μ_{ps} a média dos valores fenotípicos dos indivíduos selecionados e μ_p a média fenotípica da população.

Devido à correlação genética entre as variáveis, a seleção truncada em apenas uma variável pode levar a resultados indesejáveis nas demais. Com o objetivo de avaliar as mudanças ocorridas na outra variável, dada a seleção truncada, foram calculados os ganhos indiretos pela expressão abaixo:

$$\Delta G_{Y(X)} = p \cdot DS_{Y(X)} \cdot h_Y^2,$$

em que:

$\Delta G_{Y(X)}$ é o ganho indireto na característica secundária, em função da seleção na principal;

$DS_{Y(X)}$ é o diferencial de seleção indireto; e

h_Y^2 a herdabilidade em sentido restrito do caráter secundário.

As correlações genóticas entre duas características de interesse podem ser calculadas pela expressão:

$$\hat{\rho}_G = \text{Cov}_G(X, Y) / \sqrt{\hat{\sigma}_{g_x}^2 \hat{\sigma}_{g_y}^2}$$

sendo $\text{Cov}_G(X, Y)$ a covariância entre as duas características.

Os ganhos de seleção foram calculados com base nos valores genéticos estimados pela metodologia de quadrados mínimos e pelos valores preditos pela metodologia REML/BLUP. Para cada ciclo, foram selecionados os 20 indivíduos superiores de acordo com os valores genéticos aditivos preditos, o que equivale à seleção das melhores progênies de acordo com as médias genóticas estimadas.

3.4. Aplicação a dados experimentais

Foram utilizados dados de três ciclos de seleção recorrente da população de milho-pipoca Viçosa, do Programa de Melhoramento do Setor de Genética, do Departamento de Biologia Geral da Universidade Federal de Viçosa.

Os dados foram obtidos em experimentos de avaliação de 196 famílias de meios-irmãos no campo experimental da Universidade Federal de Viçosa em Coimbra, Minas Gerais, nos anos agrícolas de 99/00, 01/02 e 03/04. Cada experimento correspondeu a um látice simples 14 x 14. Cada parcela correspondeu a uma fileira com 5 metros, e o espaçamento utilizado entre fileiras foi de 0,9 metro. Foram semeadas 50 sementes por fileira e, após o desabaste, foram deixadas 25 plantas, totalizando uma densidade aproximada de 55.555 plantas/ha.

Em cada parcela dos experimentos, foram avaliadas a produção em kg/ha e a capacidade de expansão (CE), medida em ml/g. Os dados de produção de grãos foram corrigidos para a umidade padrão de 14,5%, em

base úmida, antes das análises estatísticas, utilizando-se a seguinte expressão:

$$PC_c = PC \cdot \frac{(100 - U)}{85,5},$$

sendo PC o peso de campo e U a umidade (%).

Os dados de produção também foram corrigidos para o estande de 25 plantas por parcela, utilizando-se o método da covariância proposto por VENCOVSKY e CRUZ (1991), de acordo com a seguinte expressão:

$$Z_{ij} = Y_{ij} - b(ST - 25),$$

em que:

Z_{ij} é o rendimento corrigido da j-ésima repetição do tratamento i;

Y_{ij} é o rendimento do tratamento i na repetição j; e

b é o coeficiente de regressão residual de Y_{ij} , em função do estande final (ST).

O coeficiente b é obtido pela seguinte expressão:

$$b = \frac{Cov_r(PG, ST)}{V_r(ST)},$$

sendo $Cov_r(PG, ST)$ a covariância residual entre peso de grãos e estande final e $V_r(ST)$ a variância residual da característica estande final.

Na determinação da CE, foram retiradas amostras de 30 gramas de grãos de cada parcela. Para o pipocamento, foi utilizada uma pipoqueira de ar quente com 1200 watts de potência. O sistema empregado foi o mesmo adotado por Matta e Viana (1996), o qual consiste em colocar os grãos no recipiente da pipoqueira quando a temperatura atinge 100 °C. Esse processo de pré-aquecimento é recomendado pelos melhoristas de milho-pipoca com o objetivo de reduzir a variação existente entre as mensurações.

Para cada ciclo de melhoramento, foram estimados os componentes de variância e preditos os valores genéticos, de acordo com a metodologia REML/BLUP descrita anteriormente, e estimados os componentes de variância e as médias genotípicas por OLS.

4. RESULTADOS E DISCUSSÃO

Como os estimadores de máxima verossimilhança seguem a distribuição normal, o teste adequado para avaliar a variabilidade genotípica em experimentos avaliados pela metodologia REML/BLUP é o teste z, diferentemente do que é feito em análises de mínimos quadrados, em que é utilizado o teste F. Essa análise da variabilidade é de fácil execução, já que as saídas do *proc mixed* fornecem as estimativas dos componentes de variância genotípica e residual com seus respectivos desvios padrão e valores de z que testam a hipótese de os componentes de variância serem iguais a zero, com sua probabilidade.

Analisando as variáveis CE e a produção nos três ciclos de seleção pôde-se observar que no primeiro ciclo de seleção foi verificada variabilidade genética tanto para CE, quanto para produção (Quadro 2). Nos demais ciclos de seleção, ocorreu uma redução significativa da variabilidade para CE, no terceiro ciclo não foi verificada variabilidade genética para essa característica. Para produção, a redução na variabilidade ocorreu de forma mais branda que para CE. Essa falta de variabilidade dificulta a continuação do processo de melhoramento por meio de novos ciclos de seleção. Viana (2007), trabalhando com famílias de meios-irmãos da população Beija-Flor de milho-pipoca, também detectou redução de variabilidade ao longo de quatro ciclos de seleção, mas de forma menos drástica, já que somente no quarto ciclo de seleção não foi verificada variabilidade genética pelo teste F.

A baixa variabilidade verificada para CE no segundo e terceiro ciclos de seleção, pode ser explicada pela alta intensidade de seleção aplicada. Avaliando a seleção direta para CE, a seleção de apenas 10% dos melhores indivíduos em cada ciclo fez com que as 20 famílias selecionadas no segundo ciclo fossem originárias de apenas 10 famílias do primeiro ciclo de seleção. Entre as 20 famílias selecionadas no terceiro ciclo, a redução é mais drástica, sendo que essas famílias se originaram apenas de oito famílias do primeiro ciclo.

Quadro 2 - Parâmetros genéticos estimados por REML, média dos tratamentos, média dos selecionados, ganhos de seleção direta e indireta para CE e produção

Parâmetros	Ciclo 0		Ciclo 1		Ciclo 2	
	CE	Produção	CE	Produção	CE	Produção
Var. genotípica	2,0415 ^{0,0146}	119281 ^{0,0130}	1,0072 ^{0,1482}	69629 ^{0,0006}	0	46085 ^{0,3057}
	± 1,4070	± 78860	± 0,6966	± 20341		± 44993
Var. aditiva	8,166	477124	4,0288	278516	0	184340
Var. residual	15,0014	981052	8,1621	187987	6,816	563160
Var. fenotípica	10,1798	651506,8990	5,7087	171612,9420	3,6981	351602,3
Herdabilidade	0,2005	0,1830	0,2243	0,4057	0	0,1311
Correlação genética	-0,306		-0,721		-	
Média	26,73	3870,26	30,72	2078,95	31,84	3730,61
Média Selecionados	27,82	4134,99	31,44	2350,90	32,57	3882,96
GS direto	0,22	48,47	0,16	110,34	0	19,97
GS indireto	0,04	-2,47	0,001	-10,3115	-0,001	0
Média IAC 112	30,30	4458,84	34,38	1600,43	29,75	3603,22

Embora tenha sido detectada essa redução de variabilidade ao longo dos ciclos de seleção, deve-se destacar que houve melhoramento genético ao longo dos ciclos. Comparando-se os valores médios de CE com a testemunha IAC 112, verifica-se um ganho de aproximadamente 6 unidades do primeiro para o terceiro ciclo de seleção, já que a média do primeiro para o terceiro ciclo de seleção passou de 26,73 para 31,63 mL/g. Enquanto isso, a testemunha obteve desempenho semelhante nos dois ensaios, em torno de 30 mL/g. Comparando o desempenho das famílias testadas com os resultados obtidos por Pacheco *et al.* (1998), o bom desempenho das populações fica mais claro. Tais autores, analisando os dados do 2º ciclo de seleção entre progênies de meios-irmãos das populações de milho-pipoca CMS 42 e CMS 43, encontraram valores médios de CE de 15,21 e 16,85 mL/g, respectivamente. Isto mostra o potencial da população em estudo como uma promissora variedade e como fonte de extração de linhagens para produção de híbridos.

Quanto à produção, também se pode destacar o bom desempenho da população Viçosa nos três ciclos de seleção, que produziu 3.870,26, 1.727,83 e 4.148,34 kg/há em cada ciclo. Esses resultados ficam acima dos encontrados por Vendruscolo *et al.* (2001), que testaram cultivares na região Centro-Sul do Brasil. Tais autores obtiveram produtividade média de 2.100 kg/ha do cultivar com melhor desempenho. Em relação à CE, esses autores obtiveram com o melhor cultivar uma CE de 21 mL/g, bem abaixo da população Viçosa nos três ciclos de seleção.

No melhoramento genético, quando se trabalha com mais de uma característica, é importante que se conheça a correlação genética existente entre essas características, para assim otimizar o ganho em todos os caracteres de interesse. Em milho-pipoca, geralmente CE e produção possuem correlação genética negativa. Viana (2007) trabalhando com quatro ciclos de seleção entre famílias de meios-irmãos, verificou que ao longo dos ciclos havia inversão do sinal e incremento na magnitude da correlação genética entre CE e produção. Isso não foi observado neste trabalho. Na verdade, o que ocorreu foi o contrário, do primeiro para o segundo ciclo a correlação se manteve negativa com aumento no seu valor absoluto (Quadro 2). Esse aumento no valor absoluto da correlação é reflexo da

seleção direta praticada para CE, que é a principal característica de interesse no melhoramento de milho-pipoca. No terceiro ciclo, não foi possível o cálculo dessa correlação já que não houve variabilidade genética.

Analisando os ganhos de seleção preditos, observa-se que eles são de pequena magnitude para as duas características avaliadas (Quadro 2). Isso se justifica ao analisar os valores de herdabilidade, que são baixos. Outro aspecto que faz com que os ganhos preditos sejam de baixa magnitude é a pequena diferença entre a média dos selecionados e a média geral. Observando os ganhos preditos indiretos, fica claro que, apesar da correlação negativa entre CE e produção, a seleção direta para ambas as características praticamente não influenciou o desempenho da outra pois esses ganhos indiretos foram praticamente nulos.

Uma forma de avaliar ganhos genéticos preditos seria o cálculo do ganho realizado, entretanto, deve-se levar em conta que o segundo e o terceiro ciclos de seleção foram realizados levando-se em conta apenas a seleção dos indivíduos que possuíam maiores médias genótípicas estimadas por OLS. Dessa maneira, fica impraticável comparar esses ganhos preditos com os realizados ao longo dos ciclos de seleção, já que são oriundos de diferentes metodologias.

Analisando os intervalos de confiança das médias genótípicas de famílias preditas, para CE (Figuras 1, 2 e 3), observa-se que, traçando uma reta ligando os pontos, à medida que avançam os ciclos de seleção avançam, a inclinação da reta diminui e há uma tendência de a reta se tornar paralela ao eixo X. Isso evidencia a redução da variabilidade genética ao longo dos ciclos de seleção. Pode-se observar também que do primeiro para o terceiro ciclo de seleção os intervalos de confiança tiveram tendência de ficar mais estreitos. Pelas análises gráficas, também é possível notar aumento da média ao longo dos ciclos de seleção, visto que as figuras se apresentam na mesma escala.

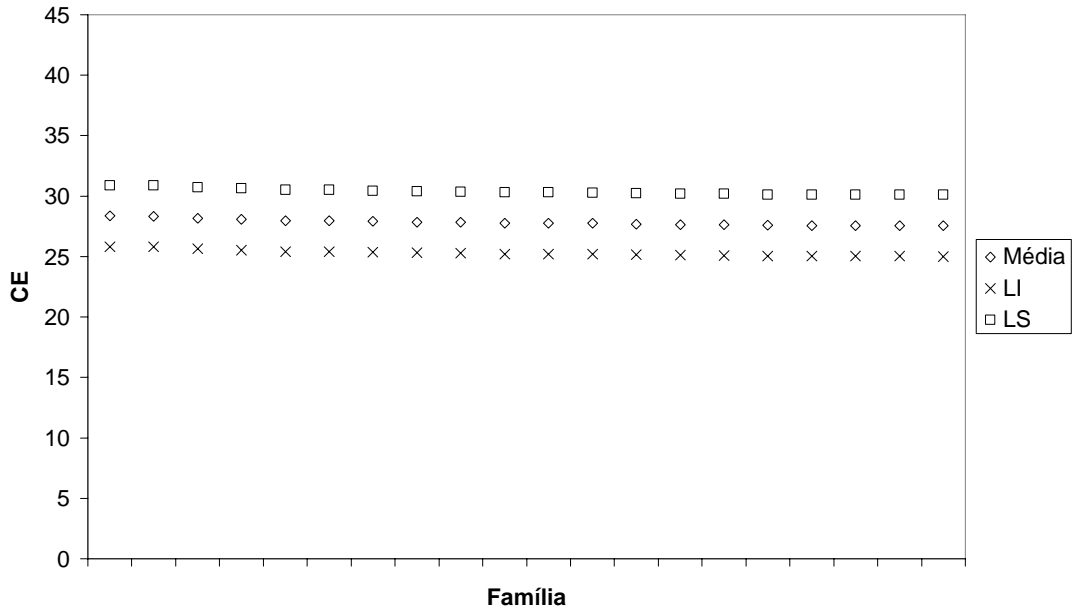


Figura 1 - Médias genotípicas previstas ($\hat{M} + 1/2 \tilde{A}$) para CE e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do primeiro ciclo de seleção.

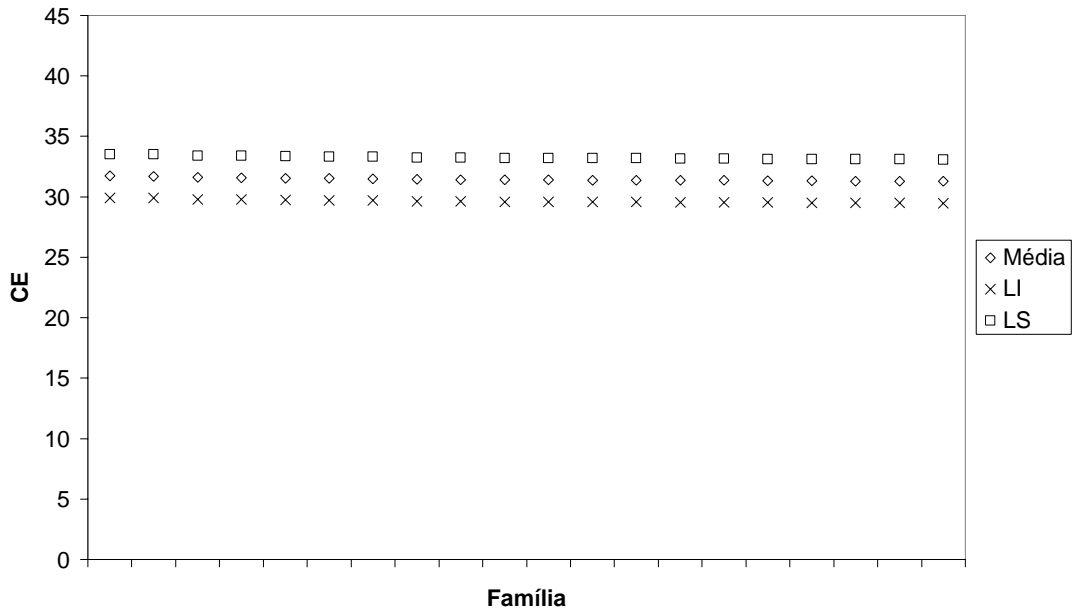


Figura 2 - Médias genotípicas previstas ($\hat{M} + 1/2 \tilde{A}$) para CE e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do segundo ciclo de seleção.

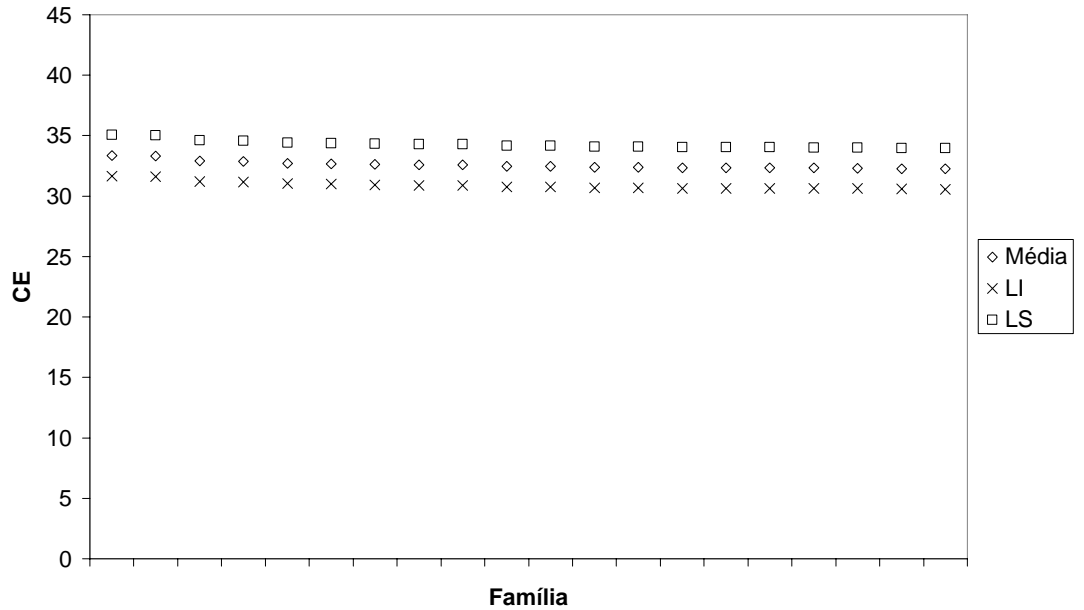


Figura 3 - Médias genotípicas preditas ($\hat{M} + 1/2 \hat{A}$) para CE e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do terceiro ciclo de seleção.

Ao tratar da produção, a situação é bastante semelhante à CE, principalmente no que se refere ao estreitamento dos intervalos ao longo dos ciclos de seleção (Figuras 4, 5 e 6). Quanto à inclinação das retas, verificou-se diminuição principalmente do segundo para o terceiro ciclo de seleção, concordando assim com o teste z, que identificou variabilidade para os dois primeiros ciclos de seleção.

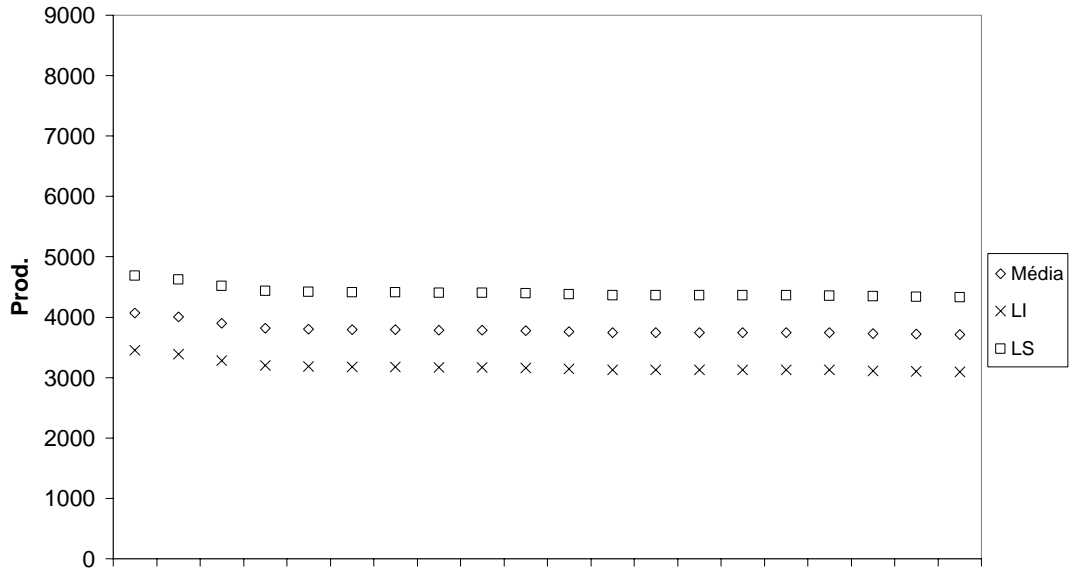


Figura 4 - Médias genotípicas preditas ($\hat{M} + 1/2 \hat{A}$) para Produção e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do primeiro ciclo de seleção.

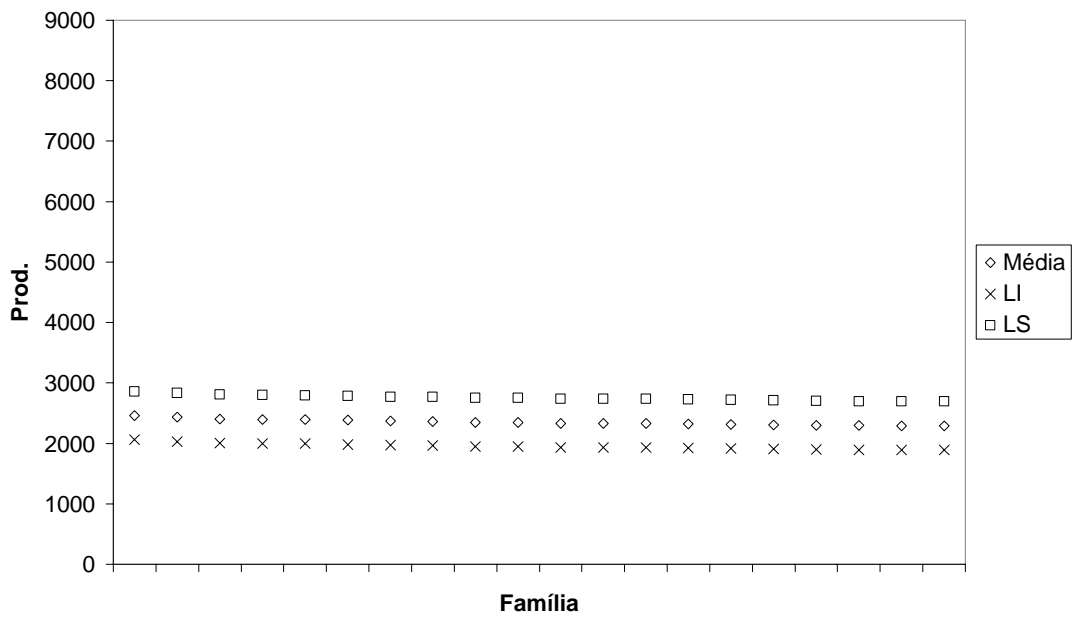


Figura 5 - Médias genotípicas preditas ($\hat{M} + 1/2 \hat{A}$) para Produção e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do segundo ciclo de seleção.

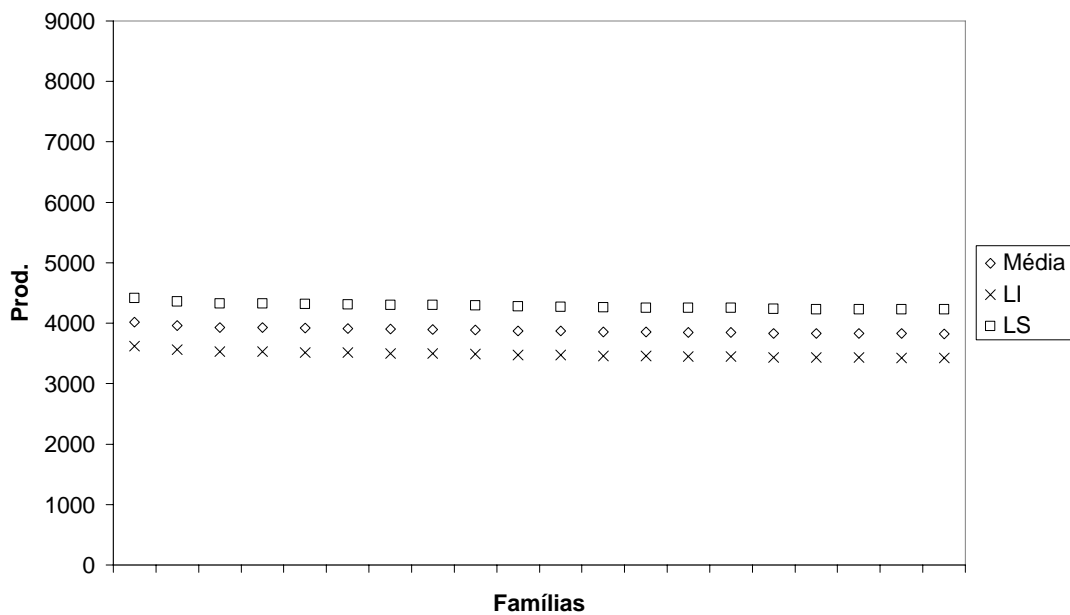


Figura 6 - Médias genotípicas preditas ($\hat{M} + 1/2 \tilde{A}$) para Produção e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do terceiro ciclo de seleção.

Ao comparar os parâmetros genéticos obtidos por REML e OLS, pode-se perceber que existe semelhança entre os valores encontrados pelas duas metodologias (Quadros 2 e 3). Entretanto, os desvios padrão para as estimativas de REML são menores que os desvios para OLS, portanto, pode-se considerar que essas estimativas de REML são mais precisas.

Quadro 3 - Parâmetros genéticos calculados por mínimos quadrados, média dos tratamentos, média dos selecionados, ganhos de seleção direta e indireta para CE e produção

Parâmetros	Ciclo 0		Ciclo 1		Ciclo 2	
	CE	Produção	CE	Produção	CE	Produção
Var. genotípica	2,8613 ^{0,0165}	113760,1 ^{0,0964}	0,8672 ^{0,1245}	73058,86 ^{0,001}	0	292792,4 ^{0,0044}
	± 1,3351	± 86320,37	± 0,743068	± 23907,54		± 112041,2
Var. aditiva	11,4452	455040,2	3,469026	292235,4	0	1171170
Var. residual	14,1673	986283,4	8,596942	230789,4	12,2218	1134465
Var. fenotípica	10,4509	642126,2	5,472761	196696	3,5625	900541,5
Herdabilidade	0,2737	0,1772	0,1585	0,3714	0	0,3251
Correlação genética	-0,195		-0,812		-	
Média	26,73	3870,26	30,72	2078,95	31,84	3730,61
Média Selecionados	32,41	5336,26	34,57	2498,54	35,78	4933,29
GS direto	1,56	244,02	0,61	255,35	0	391,03
GS indireto	0,16	-28,99	-1,51	-0,17	0,25	0
Média IAC 112	30,30	4458,84	34,38	1600,43	29,75	3603,22

Observando os valores de herdabilidade, percebeu-se pequena superioridade para os valores estimados por REML (Quadros 2 e 3). Isso aconteceu porque os valores da variância residual na metodologia REML foram menores que para OLS. Quanto aos ganhos devidos à seleção, os valores encontrados para estimativas por OLS são superiores aos encontrados por BLUP/REML. Esse fato se deve à superestimação dos valores genéticos quando se utiliza OLS. Essa superestimação foi relatada por Resende et al. (1996), que comparavam diferentes métodos de estimação/predição de valores genéticos para *Pinus*. Segundo os autores, isto ocorre porque este procedimento não considera o parâmetro herdabilidade e também é inflacionado pelos efeitos fixos de experimento.

Em relação às médias genóticas estimadas por OLS, pôde-se perceber que os valores para os indivíduos com melhor desempenho, quando comparados com as médias preditas, foram superestimados (Figuras 7 a 12). Por estas figuras, fica claro também que os intervalos de confiança para os valores estimados por OLS são maiores que os intervalos para os valores preditos por REML/BLUP. Isso demonstra o quanto as informações das médias genóticas preditas são mais confiáveis que das médias estimadas, evidenciando assim a superioridade da metodologia REML/BLUP sobre as metodologias de OLS.

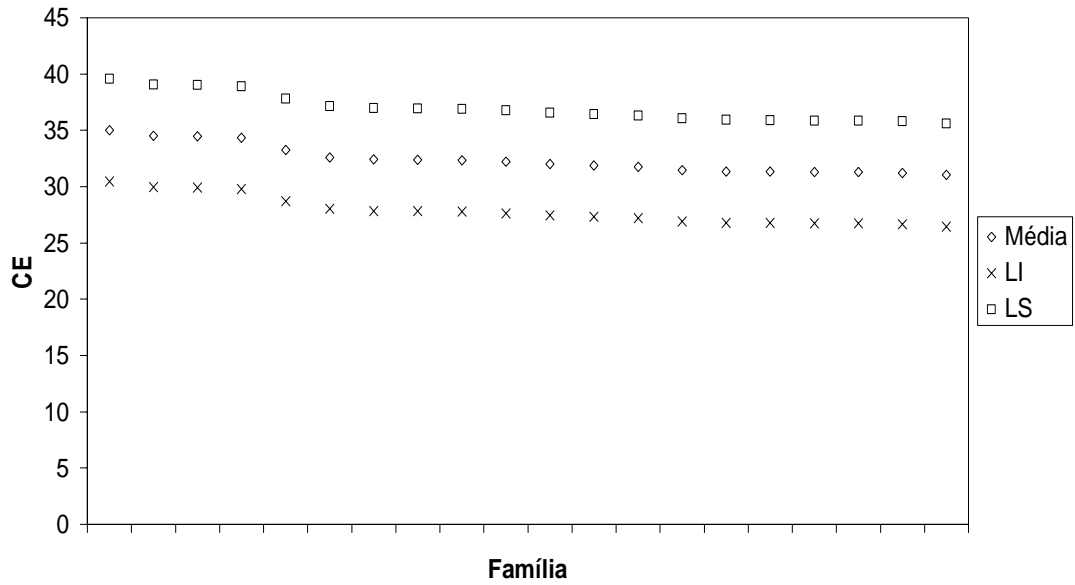


Figura 7 - Médias genótípicas estimadas ($M+1/2\hat{A}$) para CE e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do primeiro ciclo de seleção.

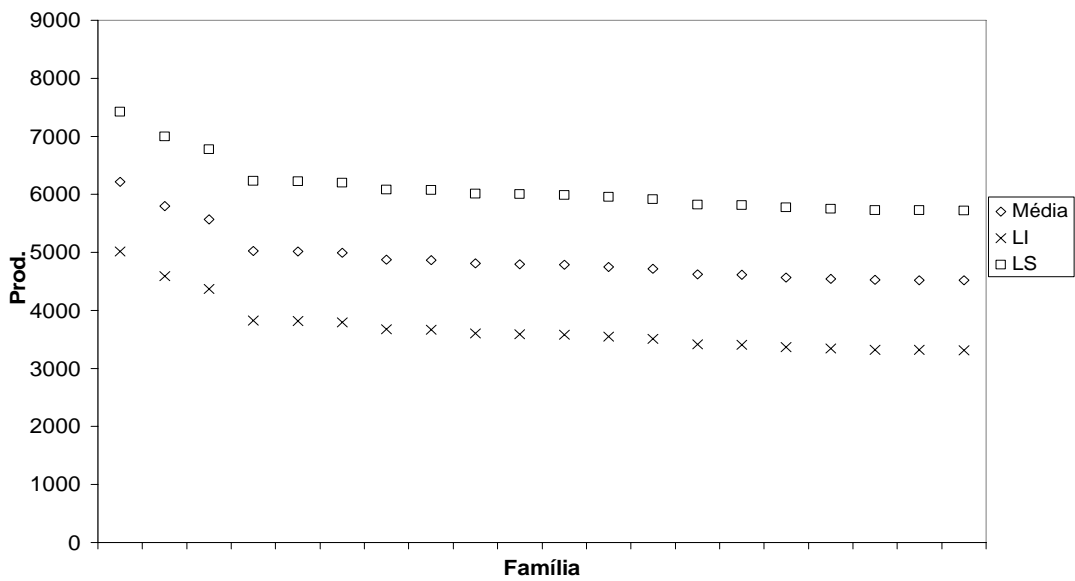


Figura 8 - Médias genótípicas estimadas ($\hat{M} + 1/2\hat{A}$) para produção e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do primeiro ciclo de seleção.

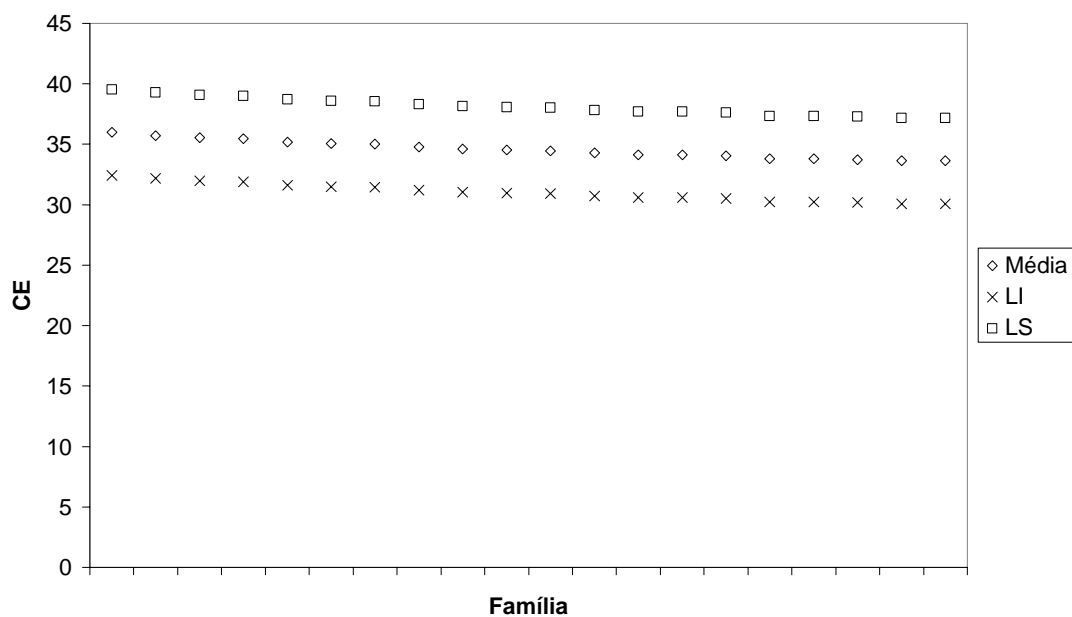


Figura 9 - Médias genotípicas estimadas ($\hat{M} + 1/2\hat{A}$) para CE e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do segundo ciclo de seleção.

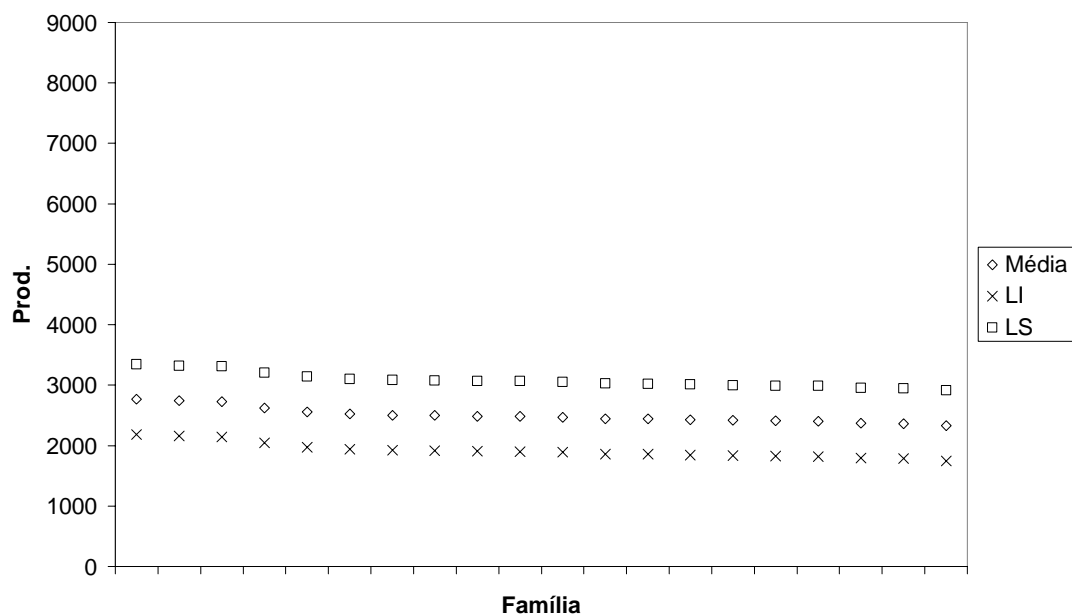


Figura 10 - Médias genotípicas estimadas ($\hat{M} + 1/2\hat{A}$) para produção e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do segundo ciclo de seleção.

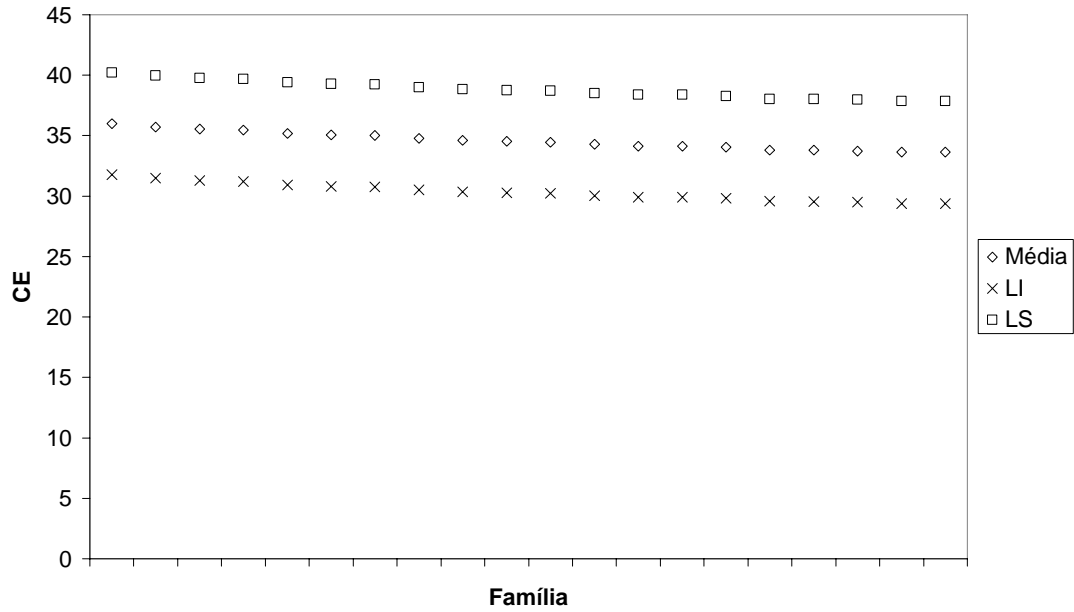


Figura 11 - Médias genótípicas estimadas ($\hat{M} + 1/2\hat{A}$) para CE e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do terceiro ciclo de seleção.

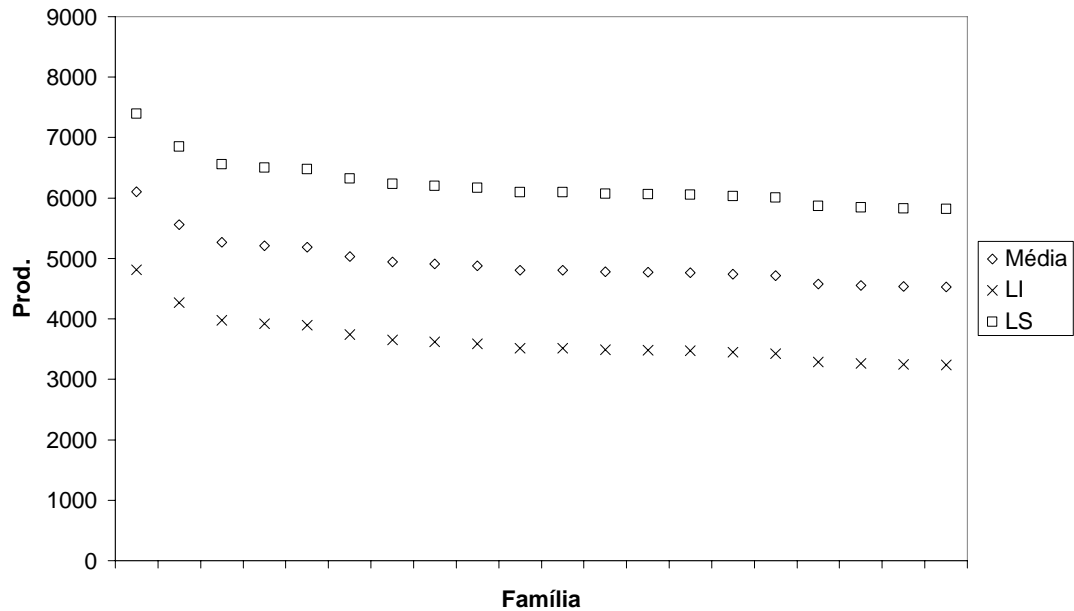


Figura 12 - Médias genótípicas estimadas ($\hat{M} + 1/2\hat{A}$) para produção e seus respectivos limites superiores (LS) e limites inferiores (LI) do intervalo de confiança de 95% para as 20 famílias de meios-irmãos selecionadas no teste do terceiro ciclo de seleção.

Em relação à ordenação dos indivíduos de acordo com os valores de médias genotípicas preditas e estimadas, para o primeiro ciclo de seleção houve grande concordância entre as duas metodologias. Esse fato é esperado, pois no primeiro ciclo não existe informação de parentesco entre os indivíduos testado, o que torna as duas metodologias semelhantes. Isso pode ser percebido pela correlação de Spearman e pelos valores de coincidência entre as famílias selecionadas pelas duas metodologias (Quadro 4). À medida que se adiciona a informação de parentesco nas análises de BLUP percebe-se que há uma redução tanto nos valores de correlação quanto nos de coincidência.

Quadro 4 - Correlação de Spearman entre médias genotípicas estimadas e preditas, coincidência entre famílias selecionadas (%)

Parâmetros	Ciclo 0		Ciclo 1		Ciclo 2	
	CE	Produção	CE	Produção	CE	Produção
Correlação de Spearman	0,98**	0,98**	0,88**	0,93**	0,84**	0,79**
Coincidência	85	95	75	80	70	75

** significativamente diferente de zero a 1% de probabilidade.

Essa diferença na ordenação dos melhores indivíduos é de grande importância no melhoramento recorrente. Como no plantio do lote de recombinação são considerados os indivíduos com maiores médias genotípicas preditas ou estimadas, fica evidente que em cada metodologia diferentes famílias seriam avançadas para a recombinação. Ao considerar os valores preditos mais precisos que os valores estimados, conclui-se que o BLUP, por meio das informações de parentesco, é a melhor alternativa de avaliação de famílias na seleção recorrente, pois, com maior precisão na identificação das famílias superiores, consegue-se um maior ganho a cada ciclo de seleção.

5. RESUMO E CONCLUSÕES

O presente trabalho teve como objetivos a implementação da metodologia REML/BLUP no melhoramento genético empregando famílias de meios-irmãos, comparar essa metodologia aos resultados obtidos com a análise tradicional de mínimos quadrados e apresentar as rotinas do SAS para análises de modelos mistos.

Os aspectos teóricos foram desenvolvidos baseados no conhecimento de genética quantitativa de acordo com abordagens de modelos mistos aplicados ao melhoramento populacional recorrente utilizando famílias de meios-irmãos. As análises computacionais foram realizadas utilizando o aplicativo computacional SAS[®], por meio dos seus procedimentos *proc mixed*, *proc inbreed*, *proc iml* e *proc glm*.

Como suporte à teoria, foram utilizado dados de três ciclos de seleção recorrente da população de milho-pipoca *Viçosa*, do Programa de Melhoramento do Setor de Genética, do Departamento de Biologia Geral da Universidade Federal de Viçosa. Cada experimento se constituiu de um látice simples de 14 x 14. Em cada parcela dos experimentos, foram avaliadas a produção em kg/ha e a capacidade de expansão (CE), medida em ml/g.

Para cada ciclo de melhoramento, foram estimados os componentes de variância e preditos os valores genéticos, de acordo com a metodologia REML/BLUP. Para cada ciclo, também foi feita análise dos dados pelos Mínimos Quadrados Ordinários (OLS), com estimação dos componentes de variância e dos valores genéticos. Foram calculados ainda os ganhos de seleção para cada ciclo, de acordo com os 20 melhores indivíduos identificados por cada metodologia. Também foi verificada a ordenação dos indivíduos selecionados e construídos os intervalos de confiança para os valores genéticos preditos e estimados.

Pela análise de variância, pôde-se verificar que à medida que os ciclos de melhoramento avançaram, houve significativa redução na variabilidade para CE, e no terceiro ciclo de seleção não foi detectada variabilidade. Essa redução de variabilidade pode ser explicada pela alta

intensidade de seleção, tanto que as 20 famílias selecionadas no terceiro ciclo eram originárias de apenas nove famílias do primeiro ciclo de seleção.

Quanto aos parâmetros genéticos estimados pelas duas metodologias observou-se que, tanto pela metodologia REML quanto por OLS, os valores estimados para os componentes de variância foram semelhantes. Para as médias genotípicas, quando não existia parentesco entre os indivíduos, houve concordância entre os melhores indivíduos estimados e preditos, porém, os valores preditos apresentaram intervalos de confiança mais estreitos, indicando que esses valores são mais precisos que os estimados. Ao incluir as informações de parentesco, houve uma redução nos valores de coincidência entre selecionados e da correlação de Spearman, indicando que o uso das médias genotípicas preditas na seleção dos indivíduos superiores proporcionará maiores ganhos devidos à seleção.

Como conclusão, fica claro que a metodologia de modelos mistos pode ser empregada no melhoramento genético, utilizando famílias de meios-irmãos desde que se tenha o controle da genealogia, já que as ferramentas para sua implementação estão disponíveis e são de fácil manuseio. Além disso, essa metodologia gera valores mais confiáveis que as metodologias tradicionais.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDRE, C. M. G. **Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores na análise dialélica.** Dissertação e Mestrado, UFLA, Lavras, 1999. 101 p.
- ANDERSON, R. D. Variance components. In: **USE of mixed model for prediction and for estimation of (co)variance components.** Armidale: University of New England, AGBU, 1984. p. 77-145.
- BAKER, R. L.; MORRIS, C. A.; JOHNSON, D. L. *et al.* Results of selection for yearling or 18-month weight in Angus and Hereford cattle. **Livestock Production Science**, v. 29, p. 277-296, 1991.
- BARBIN, D. **Componentes de variância: teoria e aplicações.** 2.ed. revisada e ampliada, Piracicaba, 1993. 120 p.: il.
- BAUER, A. M.; REETZ, T. C.; LEON, J. Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. **Crop Science**, v. 46, n. 6, p. 2685-2691, 2006.
- BAUER, A. M.; LEON, J. Multiple-trait breeding values for parental selection in self-pollinating crops. **Theor. Appl. Genet.**, v. 116, p. 235-242, 2008.
- BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v. 34, n. 1, p. 20-25, 1994.
- BERNARDO, R. Best linear unbiased prediction maize single-cross performance. **Crop Science**, Madison, v. 35, p. 141-147, 1995.
- BERNARDO, R. Genetic models for predicting maize performance in unbalanced yield trial data. **Crop Science**, Madison, v.36, p.50-56, 1996.
- BLASCO, A. The Bayesian controversy in animal breeding. **Journal Animal Science**, Champaign, v. 79, p. 2023-46, 2001.
- BOLDMAN, K. G.; KRIESE, L. A.; Van VLECK, L. D. *et al.* **A manual for use of MTDFREML: a set of programs to obtain estimates of variances and covariances (DRAFT).** Lincoln: Department of Agriculture/Agricultural Research Service, 1993. 120 p.
- CAMPOS, H.; COOPER, M.; EDMEADES, G. O.; LOFFLER, C.; SCHUSSLER, J. R.; IBANEZ, M. Changes in drought tolerance in maize associated with fifty years of breeding for yield in the US corn belt. **Maydica**, v. 52, n. 2, p. 369-381, 2006.
- CHARCOSSET, A.; BONNISSEAU, B.; TOUCHEBEUF, O.; BURSTIN, J.; BARRIÈRE, Y.; GALLAIS A.; DENIS, J.B. Prediction of Maize Hybrid

- Silage Performance Using Marker Data: Comparison of Several Models for Specific Combining Ability. **Crop Science**, v. 38, p. 38-44, 1998.
- COCHRAN, W. G.; COX, D. F. **Desenos experimentales**. Mexico: Trillas. 1978. 661 p.
- COIMBRA, J. L. M.; KOPP, M. M.; SOUSA, V. Q.; BENIN, G.; MARCHIORO, V. S.; CARVALHO, F. I. F.; OLIVEIRA, A. C. Prediction of genetic value in F3 populations of *Avena sativa* L. using Reml/Blup. **Crop Breeding and Applied Biotechnology**, v. 5, n. 3, p. 265-271, 2005.
- DUARTE, J. B.; VENCOSKY, R. Estimaco e predico por modelo linear misto com ênfase na ordenaco de mdias de tratamentos genticos **Scientia Agricola**, v. 58, n. 1, p. 109-117, jan./mar. 2001
- EBERHART, S. A. Factors effecting efficiencies of breeding methods. **African soils**, n. 15, p. 669-680, 1970.
- FALCONER, D. S.; MACKAY, T. F. C. **Introduction to Quantitative Genetics**. Fourth edition, Longmann & Co, London. 1996.
- FARIAS NETO, J. T.; RESENDE, M. D. V. Application of the mixed model methodology (REML/BLUP) in variance components estimation and prediction of genetic values in peach palm (*Bactris gasipaes*). **Revista Brasileira de Fruticultura**, Jaboticabal, v. 23, n. 2, 2001.
- FARIAS NETO, J. T.; RESENDE, M. D. V.; OLIVEIRA, M. S. P.; SANTOS, N. S. A.; CANUTO, E. L.; NOGUEIRA, O. L.; MULER, A. A. Avaliaco gentica de prognies de polinizaco aberta de aa e estimativas de parâmetros genticos. **Cerne**, Lavras, v. 13, n. 4, p. 376-383, out./dez. 2007.
- FISHER, D. S. The correlation between relatives on the supposition of Mendelian inheritance. **Review Society Transation**, Edinburgh, v. 52, p. 399-433, 1918.
- FREEMAN, A. E. Components of variance: their history, use and problems in animal breeding. In: VAN VLECK, L. D.; SEARLE, S. R., (ed). **Variance components and animal breeding**. Ithaca, Cornell Univerity, 1979.
- FREITAS, A. R. Avaliaco de procedimentos da estimaco de parâmetros genticos em bovinos de corte. **Revista Brasileira Zootecnia**, v. 29, n. 1, p. 94-102, 2000.
- FREITAS, A. R. **Comparaco de métodos de estimaco de componentes de variâncias e parâmetros afins de múltiplos caracteres em bovinos**. Piracicaba: ESALQ/USP, 1991. 170 p. Tese Doutorado

- GARCIA, C. H.; NOGUEIRA, M. C. S. N. Utilização da metodologia REML/BLUP na seleção de clones de eucalipto. **Scientia Forestalis**, n. 68, p. 107-112, 2005.
- GIANOLA, D.; FERNANDO R. L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v. 63, p. 217-244, 1986.
- GIANOLA, D.; FOULLEY, J. L.; FERNANDO, R. L. *et al.* Estimation of Heterogeneous Variances Using Empirical Bayes Methods: Theoretical Considerations. **Journal of Dairy Science**, v. 75, p. 2805-2823, 1992.
- GONÇALVES, E. S. T.; AUBYN, A.; MARTINS, A. Mixed spatial models for data analysis of yield on large grapevine selection field trials. **Theoretical and Applied Genetics**, v. 115, n. 5, p. 653-663, 2007.
- HARTLEY, H. O.; RAO, J. N. K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**, v. 54, p. 93-108, 1967.
- HENDERSON, C. R. Estimation of genetic parameters. **Ann. Math. Stat.**, (Abstra.), v. 21, p. 309, 1950.
- HENDERSON, C. R. Estimation of variance and covariance components. **Biometrics**, Washington, v. 9, p. 226-252, 1953.
- HENDERSON, C. R. General flexibility of linear model techniques for sire evaluation. **Journal of Dairy Science**, v. 57, p. 963, 1974.
- HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462 p.
- HENDERSON, C. R. Best linear unbiased prediction using relationship matrices derived from selected base populations. **Journal of Dairy Science**, v. 68, n. 2, p. 443-448, 1985.
- HENDERSON, C. R. Between algorithms: A "Short Cut" restricted maximum likelihood procedure to estimate variance components. **Journal of Dairy Science**, v. 69, n. 7, p. 1904-1916, 1986.
- IEMMA M. **Uso do melhor preditor linear não viesado (BLUP) em análises dialélicas e predição de híbridos**. Dissertação de Mestrado, ESALQ, USP, Piracicaba, 2003. 81 p.
- KALIL FILHO, A. N.; RESENDE, M. D. V.; KALIL, G. P. C. Componentes de variância e predição de valores genéticos em seringueira pela metodologia de modelos mistos (REML/BLUP). **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 9, p. 1883-1887, 2000.
- KENNEDY, B. W. C. R. Henderson: the unfinished legacy. **Journal of Dairy Science**, v. 74, p. 4067-4081, 1991.

- LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS System for Mixed Models**, 2.ed., Cary, NC: SAS Institute Inc., 2006.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland, Massachusetts, USA: Sinauer Associates, Inc. 980 p. 1998.
- MATTA, F. P.; VIANA, J. M. S. **Heterose e conseqüências de endogamia em populações de milho pipoca (*Zea mays* L.)**. In: Simpósio de Iniciação Científica, 6, Viçosa, 1996. Resumos. Viçosa, Imprensa Universitária, 1996. p. 33.
- MEYER, K. Between algorithms: A "Short Cut" restricted maximum likelihood procedure to estimate variance components. **Journal of Dairy Science**, v. 69, n. 7, p.1904-1916, 1986.
- MORENO-GONZALES, J.; ARES, J. L. A.; FERRO, R. A.; RAMIREZ, L. C. Genetic and statistical models for estimating genetic parameters of maize seedling resistance to *Fusarium graminearum* Schwabe root rot. **Euphytica**, v. 137, n. 1, p. 55-61, 2004.
- MRODE, R. A. **Linear models for the prediction of animal breeding values**. 2.ed. 2005. 344 p.
- NAOE, L. K.; SEDIYAMA, C. S.; PIOVESAN, N. D.; LOPES, P. S.; COIMBRA, R. C.; BARROS, E. G.; MOREIRA, M. A. Estimação de componentes de variância e parâmetros genéticos considerando o parentesco em população de retrocruzamento de soja. **Revista Ciência Agroambiental**, v. 1, n. 1, janeiro a junho de 2006.
- NUNES, J. A. R. **Incorporação da informação de parentesco no método genealógico pelo enfoque de modelos mistos**. 113 f. Tese (Doutorado em Agronomia - Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG, 2006.
- PACHECO, C. A. P.; GAMA, E. E. G.; GUIMARÃES, P. E. O.; SANTOS, M. X.; FERREIRA, A. S. Estimativas de parâmetros genéticos nas populações CMS-42 e CMS-43 de milho-pipoca. **Pesquisa Agropecuária Brasileira**, v. 33, n. 12, p. 1995-2001, 1998.
- PARNELL, P. F.; ARTHUR, P. F.; BARLOW, R. Direct response to divergent selection for yearling growth rate in Angus cattle. **Livestock Production Science**, v. 49, p. 297- 304, 1997.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when blocks sizes are unequal. **Biometrika**, v. 58, p. 545-554, 1971.

- PIEPHO, H. P.; MOHRING, J.; MELCHINGER, A. E.; BUCHSE, A. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, p. 209-228, 2008.
- PERRI, S. H. V.; IEMMA, A. F. Procedure of software SAS® for the analysis of mixed models. **Scientia agricola**, Piracicaba, v. 56, n. 4, 1999.
- RAZOOK, A. G.; FIGUEIREDO, L. A.; BONILHA NETO, L. M. *et al.* Selection for yearling weight in Nelore and Guzera zebu breeds: selection applied and response in 15 years of progeny. (compact disc). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. **Proceedings...** Armidale, 1998. CD-ROM. Session: Beef Cattle.
- RESENDE, M. D. V.; PRATES, D. F.; JESUS, A.; YAMADA, C. K. Estimaco de componentes de varincia e predico de valores genticos pelo mtodo da mxima verossimilhana restrita (REML) e melhor predico linear no viciada (BLUP) em Pinus. **Boletim de Pesquisa Florestal**, v. 32/33, p. 18-45, 1996.
- RESENDE, M. D. V. **Avanos da biomtrica florestal**. In: ENCONTRO SOBRE TEMAS DE GENTICA E MELHORAMENTO, 1997, Piracicaba: Escola Superior de Agricultura "Luiz de Queiroz".
- RESENDE, M. D. V.; FURLANI-JNIOR, E.; MORAES, M. L. T.; FAZUOLI, L. C. Estimativas de parmetros genticos e predico de valores genotpicos no melhoramento do cafeeiro pelo procedimento REML/BLUP **Bragantia**, Campinas, v. 60, n. 3, p.185-193, 2001.
- RESENDE, M. D. V. **Gentica biomtrica e estatstica no melhoramento de plantas perenes**. Braslia: EMBRAPA Informaco tecnolgica, 2002. 975 p.
- ROCHA, M. G. B.; PIRES, I. E.; XAVIER, A.; CRUZ, C. D.; ROCHA, R. B. Avaliaco gentica de prognies de meios-irmaos de *Eucalyptus urophylla* utilizando os procedimentos REML/BLUP E E(QM) **Cincia Florestal**, v. 16, n. 4, p. 369-379, 2006.
- ROTHSCHILD, M. F.; HENDERSON, C. R.; QUAAS, R. L. Effects of selection on variances and covariances of simulated first and second lactations. **Journal of Dairy Science**, v. 62, p. 996-1002, 1979.
- SAS INSTITUTE. 2000. **SAS/STAT User's guide: statistics**, verso 8.01, v. 2, 4.ed., Cary
- SCHEFF, H. **Analysis of variance**. London: John Wiley & Sons, 1959. 467 p.
- SEARLE, S. R.; CASELLA, G.; Mc CULLOCH, C. E. **Variance components**. New York: John & Wiley, 1992. 528 p.

- SEBOLAI, B.; PEDERSEN, J. F.; MARX, D. B.; BOYKIN, D. L. Effect of control plot density, control plot arrangement, and assumption of random or fixed effects on nonreplicated experiments for germplasm screening using spatial models. **Crop Science**, v. 45, n. 5, p. 1978-1984, 2005.
- SMITH, A. B.; LIW, P.; CULLIS, B. R. The design and analysis of multi-phase plant breeding experiments. **Journal of Agricultural Science**, v. 144, n. 5, p. 393-409, 2006.
- THOMPSON, R. Sire evaluation. **Biometrics**, Washington, v. 35, p. 339-353, 1979.
- VENCOVSKY, R.; CRUZ, C. D. Comparação de métodos de correção do rendimento de parcelas com estandes variados. I. Dados simulados. **Pesquisa Agropecuária Brasileira**, v. 26, n. 5, p. 647-657, 1991.
- VENDRUSCOLO, E. C. G.; SCAPIN, C. A.; PACHECO, C. A. P.; OLIVEIRA, V. R.; BRACCINI, A. L.; VIDIGAL, M. C. G. Adaptabilidade e estabilidade de cultivares de milho-pipoca na região centro-sul do Brasil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 36, n. 1, p. 123-30, 2001.
- VIANA, J. M. S.; REGAZZI, A. J. Estimation of genetic parameters in the square lattice analysis. **Bragantia**, Campinas, v. 58, n. 1, 1999.
- VIANA, J. M. S. Relative importance of the epistatic components of genotypic variance in non-inbred populations. **Crop Breeding and Applied Biotechnology**, v. 4, p. 18-27, 2004.
- VIANA, J. M. S. Melhoramento intrapopulacional recorrente de milho-pipoca, com famílias de meios-irmãos. **Revista Brasileira de Milho e Sorgo**, v. 6, n. 2, p. 199-210, 2007.
- WALSH, J. B. **Mixed-model and bayesian analysis of shortterm selection experiments**. http://biosci.arizona.edu/zbook/volume_2/chapters/vol2_07.html. 1999.
- WHITE, T. L.; HODGE, G. R. Predicting breeding values with applications in forest tree improvement. **Dordrecht: Kluwer Academic Publishe**, 1989.

ANEXOS

Anexo 1 - Exemplo de pedigree para análise de três ciclos de seleção recorrente entre famílias de meios-irmãos maternos

Indivíduo	Pai	Mãe
1	.	.
2	.	.
3	.	.
...		
196	.	.
197	.	180
...		
392	.	192
393	.	316
...		
588	.	327
589	.	373
...		
784	.	288
785	.	744
...		
979	.	633
980	.	734

Anexo 2 - Saída do proc mixed do programa SAS®

The Mixed Procedure

Model Information

Data Set	WORK.TESTE1
Dependent Variable	CE
Covariance Structure	Linear
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Blo	28	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Rep	2	1 2

The Mixed Procedure

Class Level Information

Class	Levels	Values
Trat	196	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196

The Mixed Procedure

Dimensions

Covariance Parameters	2
Columns in X	31
Columns in Z	196
Subjects	1
Max Obs Per Subject	392
Observations Used	392
Observations Not Used	0
Total Observations	392

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	1924.97671806	
1	2	1922.77211077	0.00000003
2	1	1922.77208996	0.00000000

Convergence criteria met.

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
LIN(1)	1.0072	0.6966	1.45	0.1482
Residual	8.4553	0.8808	9.60	<.0001

Fit Statistics

-2 Res Log Likelihood	1922.8
AIC (smaller is better)	1926.8
AICC (smaller is better)	1926.8
BIC (smaller is better)	1933.3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	2.20	0.1376

The Mixed Procedure
Solution for Fixed Effects

Effect	Blo	Rep	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			30.0186	0.8187	195	36.67	<.0001
Rep		1	0.9167	1.1514	169	0.80	0.4270
Rep		2	0
Blo(Rep)	1	1	2.1013	1.1553	169	1.82	0.0707
.							
.							
Blo(Rep)	28	2	-0.1253	1.1555	169	-0.11	0.9138

Solution for Random Effects

Effect	Trat	Estimate	Std Err Pred	DF	t Value	Pr > t
Trat	1	0.7095	0.9058	169	0.78	0.4345
.						
.						
Trat	196	-0.09474	0.9064	169	-0.10	0.9169

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Rep	1	169	37.52	<.0001
Blo(Rep)	26	169	3.89	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t	
Media BLUP trat	1	31.4306	0.9119	195	34.47	<.0001
.						
.						
Media BLUP trat	196	30.6264	0.9124	195	33.57	<.0001

Estimates

Label	Lower	Upper	
Media BLUP trat	1	29.6322	33.2290
.			
.			
Media BLUP trat	196	28.8269	32.4258