Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Comparação de métodos de estimação de componentes de variância e
parâmetros genéticos considerando o Delineamento III aplicado a caracteres
quantitativos em milho

Angela Mello Coelho

Tese apresentada, para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

Piracicaba

2010

Angela Mello Coelho
Licenciada em Matemática

**Comparação de métodos de estimação de componentes de variância e
parâmetros genéticos considerando o Delineamento III aplicado a caracteres
quantitativos em milho**

Orientador:

Prof. Dr. **Décio Barbin**

Tese apresentada, para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

Piracicaba

2010

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Coelho, Angela Mello

Comparação de métodos de estimação de componentes de variância e parâmetros genéticos considerando o Delineamento III aplicado a caracteres quantitativos em milho / Angela Mello Coelho. - - Piracicaba, 2010.
101 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2010.

1. Análise de variância 2. Componentes de variância 3. Delineamento experimental
4. Genética quantitativa 5. Herdabilidade 6. Milho 7. Verossimilhança I. Título

CDD 633.15
C672c

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

AGRADECIMENTOS

Aos meus pais Sônia e Viriato pelo apoio, amor, carinho e confiança na minha capacidade de vencer desafios e finalizar os projetos que inicio.

Ao professor Décio Barbin, por me aceitar novamente como orientada, me guiando, ajudando e apoiando em mais essa etapa da vida, e por estar presente nas horas mais críticas do desenvolvimento da tese, sempre com uma palavra amiga, me passando confiança em que tudo daria certo.

Ao professores Cláudio Lopes de Souza Júnior e Roland Vencovsky do Departamento de Genética (ESALQ/USP) por toda a ajuda, sem a qual não seria possível concluir esse trabalho.

Ao CNPq pela bolsa de estudos que permitiu minha estadia em Piracicaba e minha dedicação exclusiva ao desenvolvimento da tese.

Aos meus colegas da pós graduação, principalmente ao Vanderly, Lú, Wilson, Renata, Fernanda e Símona por toda a ajuda, companheirismo e por fazerem do doutorado uma época da qual sentirei muitas saudades.

Aos professores do Departamento de Ciências Exatas (ESALQ/USP) por todo o conhecimento dividido.

Às secretárias Luciane e Solange e aos técnicos de informática Jorge e Eduardo pela amizade e por toda a ajuda em horas de sufoco.

Aos meus vizinhos, Mateus, Sanzio e Diego por toda a amizade e paciência em me ajudar na parte genética.

Aos meus irmãos: Chico por honrar seu papel de irmão mais velho oferecendo apoio, conforto, proteção e amizade; Juliana e Guilherme meus irmão mais novos, por deixarem minha vida mais colorida e cheia de vida.

As minhas queridas amigas Mafaldetes, Débora, Fernanda, Babi e Marina, muito obrigada por sempre estarem ao meu lado, me perdoadando nas épocas de loucura e mau-humor e compartilhando das minhas conquistas como se fossem suas.

Aos meus amigos da dança de salão, Daniel, Netto, Robertinha, Sueli e Giba, por terem me mantido sã nas horas de maiores dificuldades com a tese, fornecendo algumas horas de lazer e exercício físico que permitiam a minha volta aos estudos com energia

redobrada.

À Iná e Ilhoa por existirem e serem os seres vivos carinhosos e maravilhosos que são, fazendo da minha casa um lar.

SUMÁRIO

RESUMO	7
ABSTRACT	9
1 INTRODUÇÃO	11
2 REVISÃO BIBLIOGRÁFICA	13
2.1 Melhoramento Genético do Milho (<i>Zea mays</i> L.)	13
2.2 Componentes de Variância	13
2.2.1 Componentes de Variância Genética	14
2.3 Coeficiente de Herdabilidade	15
2.4 Grau Médio de Dominância	15
2.5 Delineamento Genético	16
2.6 Delineamento Experimental	20
2.7 Métodos de Estimção de Componentes de Variância	21
2.7.1 Método da Análise da Variância	21
2.7.2 Método da Máxima Verossimilhança	22
2.7.2.1 Método da Máxima Verossimilhança Restrita	22
3 MATERIAL E MÉTODOS	25
3.1 Material	25
3.1.1 Dados Reais	25
3.1.1.1 Material Genético	25
3.1.1.2 Instalação dos Experimentos	26
3.2 Métodos	27
3.2.1 Análises Estatísticas para Delineamentos em Látice Quadrado	27
3.2.1.1 Modelo Matemático	29
3.2.1.2 Estimção dos Parâmetros Aleatórios	31
3.2.1.2.1 Método da Análise da Variância	31
3.2.1.2.2 Método da Máxima Verossimilhança Restrita (REML)	37
3.2.1.2.3 Preditores dos Parâmetros Genéticos	38
3.2.1.3 Estimção dos Efeitos Fixos	39
3.3 Dados Simulados	42

3.3.0.4 Simulação dos Dados	42
3.3.1 Comparação dos Métodos de Estimação	46
3.3.1.1 Gráficos	47
3.3.1.2 Estatísticas Descritivas	47
4 RESULTADOS E DISCUSSÃO	49
4.1 Dados Reais	49
4.2 Dados Simulados	55
4.2.1 Componentes de Variância	55
4.2.2 Parâmetros genéticos	75
5 CONCLUSÕES	95
5.1 Componentes de Variância	95
5.2 Parâmetros Genéticos	96
REFERÊNCIAS	99

RESUMO

Comparação de métodos de estimação de componentes de variância e parâmetros genéticos considerando o Delineamento III aplicado a caracteres quantitativos em milho

Esse trabalho teve como objetivo comparar métodos de estimação de componentes de variância e parâmetros genéticos, considerando tanto o delineamento estatístico fatorial instalado em látice quadrado como o Delineamento III. Como referência, foram utilizados três conjuntos de dados reais, em melhoramento genético de milho, relativos aos caracteres: produção de grãos (gramas por parcela); altura da folha bandeira ao chão (em centímetros); e o número de folhas entre a primeira espiga e o pendão. O método da Análise da Variância (ANOVA), conforme indicado pelo Delineamento III, foi utilizado na análise dos dados e estimação dos componentes de variância relativos ao modelo matemático, variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância para cada um dos três caracteres estudados. Essas estimativas foram utilizadas na simulação de 1000 conjuntos de dados com características semelhantes a cada um dos conjuntos de dados reais considerados. Os métodos da ANOVA e da máxima verossimilhança restrita (REML) foram utilizados na predição dos parâmetros já mencionados para cada um dos conjuntos de dados simulados dentro de cada caráter. As 1000 estimativas obtidas por cada método, para cada caráter estudado, foram utilizadas no cálculo de estatísticas descritivas (média, desvio-padrão e acurácia relativa) e na montagem de gráficos de box-plot. Utilizando as informações obtidas a partir das estimativas fornecidas por cada método e em posse dos valores reais que essas estimativas deveriam prever (valores utilizados na simulação dos dados) foi possível comparar ambos os métodos quanto à eficiência das estimativas por eles fornecidas. Ambos os métodos apresentaram características semelhantes na predição da maioria dos componentes de variância relativos ao modelo matemático, sendo que as maiores disparidades se deram para os componentes relativos aos efeitos de progênie (σ_p^2) e para as interações entre progênie e linhagem (σ_{pt}^2) e entre progênie, linhagem e ambiente (σ_{pta}^2); os quais são os componentes de maior peso no cálculo das variâncias e parâmetros genéticos. O método da ANOVA foi muito eficiente na predição de σ_p^2 , sendo que o método da REML se aproximou dos resultados obtidos pelo método da ANOVA conforme diminuíram os valores de referência para esse componente; para σ_{pt}^2 o método da REML se mostrou mais eficiente conforme maior o valor de referência, porém, perdeu eficiência e se aproximou do método da ANOVA conforme o valor de referência do componente diminuiu. Ambos os métodos se mostraram ineficientes na predição de σ_{pta}^2 , porém o método da REML foi o menos eficiente. O melhor desempenho do método da ANOVA na predição dos componentes de variância de maior peso no cálculo das variâncias genéticas levou a um melhor desempenho desse método na predição de todos os parâmetros genéticos, com exceção da variância de dominância, a qual dependia unicamente de σ_{pt}^2 . Porém, foi observada uma tendência no método da ANOVA, em média, na superestimação do grau médio de dominância em cerca de 45% do seu valor de referência, independentemente do caráter estudado.

Palavras-chave: Análise da Variância; Máxima Verossimilhança Restrita; Delineamento III; Componentes de Variância; Coeficiente de Herdabilidade; Grau Médio de Dominância

ABSTRACT

Comparison of estimation methods for variance components and genetic parameters considering the Design III applied to quantitative characters in maize

This work aimed to compare estimation methods for variance components and genetic parameters, considering the factorial statistical design set in randomized blocks and the genetic Design III. As reference, three sets of real data were used, on maize genetic improvement, related to the characters: grain yield (grams by plot), plant height, measured from the ground to the flag leaf in centimeters, and the number of leaves above the uppermost ear. The analysis of variance method (ANOVA), accordingly to the proposed by the Design III, was used on the analysis of the data and estimation of the variance components derived from the mathematical model, genetic variances, heritability and average degree of dominance for each of the studied characters. These estimatives were used on the simulation of 1000 data sets with similar characteristics to the real data analyzed. The ANOVA and restricted maximum likelihood (REML) methods were used on the prediction of the already mentioned parameters for each of the simulated data sets within each character. The 1000 estimatives obtained by each method, for each studied character, were used on the calculation of descriptive statistics (mean, standard deviation and relative accuracy) and for the fitting of box-plot graphics. Through the information obtained from the estimatives given by each method and in possession of the actual values that they should predict (values used in the simulation of the data sets) it was possible to compare both methods as to the efficiency of the estimatives given by them. Both methods presented similar characteristics on the prediction of most of the variance components derived from the mathematical model, being that most differences were pertinent to the components related to the effects of progeny (σ_p^2) and to the interactions between progeny and parental inbred (σ_{pt}^2) and between progeny, parental inbred and environment (σ_{pta}^2); which are the components of greater importance on the calculation of the genetic parameters. The ANOVA method was very efficient on the prediction of σ_p^2 , being that the smaller the reference value for this component, more the REML method approached the results obtained by the ANOVA method; for larger values of σ_{pt}^2 the most efficient was the REML method, but its efficiency decayed and approached the ANOVA method for smaller reference values for this component. Both methods were poorly efficient on the prediction of σ_{pta}^2 , but the REML method was the least efficient. The better performance of the ANOVA method on the prediction of the variance components of greater importance on the calculation of the genetic variances lead to a better performance of the ANOVA method on the prediction of all genetic parameters, with exception to the dominance variance, which depended solely on σ_{pt}^2 . However, it was observed a tendency on the ANOVA method, in average, on the overestimation of the average degree of dominance of around 45% of the actual reference value, independently of the studied character.

Keywords: Analysis of Variance; Restricted Maximum Likelihood; Design III; Variance Components; Heritability; Average Degree of Dominance

1 INTRODUÇÃO

O melhoramento genético de plantas vem sendo utilizado pelo ser humano há cerca de 6000 anos, visando principalmente ao aumento da produção a fim de assegurar a sobrevivência da espécie (ALLARD, 1971). Nos dias de hoje o melhoramento genético de plantas continua sendo de grande importância para os seres humanos, pois através dele podem ser obtidos novos cultivares com caracteres vantajosos, como maior produtividade, maior resistência a pragas, doenças e variações no clima. Existem diversas maneiras de fazer a seleção de indivíduos a serem utilizados no melhoramento de uma espécie; têm-se a seleção natural, sem a influência do homem, e a seleção artificial, baseada na experimentação.

Foi demonstrado por Fisher¹ (1918 apud SOUZA JÚNIOR, 1989) que a covariância entre parentes é função dos componentes de variância genética (variâncias aditiva, dominante e epistática). O desdobramento da covariância entre parentes em componentes de variância genética é de grande importância no melhoramento genético, pois possibilita estudar os processos seletivos com uma base científica. Esse método torna possível estudar a estrutura genética de uma população por meio de estimativas de componentes de variância genética e, também, de parâmetros relacionados com os processos seletivos, como por exemplo o coeficiente de herdabilidade de diferentes caracteres (SOUZA JÚNIOR, 1989).

Os componentes de variância genética podem ser estimados por meio de componentes de variância no sentido geral, que são variâncias associadas aos efeitos aleatórios de um modelo matemático (BARBIN, 1993). O modelo estudado nesse trabalho é relativo a um experimento em melhoramento de milho, implantado segundo o delineamento fatorial, em látice 10x10 com duas repetições, em diferentes ambientes; além de seguir um delineamento estatístico, o experimento em questão seguiu, também, um delineamento genético, sendo este o Delineamento III, proposto, inicialmente, por Comstock e Robinson (1952).

Existem diversos métodos de estimação de componentes de variância, porém dois deles são encontrados com mais frequência em estudos realizados com milho. O método mais comum é o método da análise da variância (ANOVA); porém, em trabalhos mais recentes foi encontrada, também, a utilização do método da máxima verossimilhança restrita (REML).

¹FISHER, R.A., The correlation between relatives on the supposition of Mendelian inheritance, Edinburgh, 52, 399-433, 1918.

No entanto, não foram encontrados na literatura, trabalhos que discutissem qual dos dois métodos é o mais apropriado na estimação de componentes de variância, coeficientes de herdabilidade e de outros parâmetros de interesse para o melhoramento genético vegetal.

Esse trabalho tem como objetivo principal comparar os métodos de estimação da ANOVA e da REML com respeito à predição, tanto dos componentes de variância do modelo matemático, quanto das variâncias e parâmetros genéticos.

O processo de comparação se iniciou com a análise, pelo método da ANOVA, de três conjuntos de dados reais relativos aos caracteres: produção de grãos (gramas por parcela); altura da folha bandeira ao chão (em centímetros); e o número de folhas entre a primeira espiga e o pendão. Em seguida foram estimados, também pelo método da ANOVA, os efeitos fixos e componentes de variância do modelo matemático, as variâncias aditivas, o coeficiente de herdabilidade e o grau médio de dominância.

Os valores preditos para os efeitos fixos e componentes de variância relativos a cada um dos caracteres estudados foram utilizados na simulação de 1000 conjuntos de dados simulados com características semelhantes às dos dados reais para cada caráter. Os métodos da ANOVA e REML foram utilizados no cálculo das estimativas dos componentes de variância e parâmetros genéticos para cada um dos 1000 conjuntos de dados simulados dentro de cada um dos três caracteres estudados.

Sendo assim, foram obtidos conjuntos de 1000 estimativas (para cada método) por componente de variância e parâmetro genético dentro de cada caráter. esses conjuntos de estimativas foram utilizados no cálculo de estatísticas descritivas (média, desvio padrão e acurácia relativa) e gráficos de box-plot. Utilizando os resultados fornecidos pelas estatísticas descritivas, pelos gráficos e de posse do valor de referência que cada estimativa deveria prever, aquele utilizado na simulação dos dados, foi possível concluir o processo de comparação dos métodos de estimação da ANOVA e da REML, considerando o caso do melhoramento genético de milho, delineamentos utilizados e caracteres estudados.

2 REVISÃO BIBLIOGRÁFICA

2.1 Melhoramento Genético do Milho (*Zea mays* L.)

Originário das Américas, o milho é um dos cereais de maior importância econômica e social no mundo, se não o de maior importância. No Brasil, é o segundo mais importante (perdendo somente para a soja) considerando produção e área semeada; é plantado em todos os estados brasileiros e em todos os tipos de propriedades, da agricultura familiar às grandes exportadoras (VILARINHO, 2005). A importância do milho se deve às diversas formas em que se pode utilizá-lo, como, por exemplo, fonte de alimento humano ou animal, ou na produção industrial de adoçantes e álcool, entre outros. A principal participação do milho é na produção de ração para suínos, aves, bovinos e animais de estimação. Como a criação de animais e a demanda por alimentos de melhor qualidade estão aumentando, nada mais natural do que a demanda pelo aumento na produção de milho. Uma ferramenta importante para o aumento dessa produção é o melhoramento genético (EMBRAPA, 2009).

2.2 Componentes de Variância

Do ponto de vista genético, o estudo de um caráter quantitativo se baseia na sua variação total. A idéia central do estudo da variação de um caráter é reparti-la em componentes de variação que podem ser atribuídos a diferentes fatores. A grandeza relativa desses componentes de variância determina as propriedades genéticas da população para um determinado caráter (FALCONER, 1993). Já do ponto de vista estatístico, componentes de variância, são variâncias associadas aos efeitos aleatórios de um modelo matemático (BARBIN, 1993).

Nos estudos de herança de caracteres quantitativos é necessário mesclar as abordagens genética e estatística, pois os estudos de um caráter métrico são feitos com o auxílio de delineamentos experimentais, os quais são estudados por meio de modelos matemáticos. Mesmo que os componentes de variância encontrados ao analisar um modelo matemático não sejam exatamente os procurados pelos melhoristas, existem maneiras de calcular os componentes de variância genéticos por meio dos componentes associados a um modelo matemático.

2.2.1 Componentes de Variância Genética

O valor observado ou medido, em um indivíduo, é denominado valor fenotípico. Esse valor pode ser decomposto em dois componentes devidos às influências do genótipo do indivíduo e do ambiente no qual ele se desenvolveu. Por genótipo entende-se a combinação de genes do indivíduo, e por ambiente entende-se qualquer outro fator não genético que possa influenciar o fenótipo (FALCONER, 1993).

Fisher¹ (1918 apud SOUZA JÚNIOR, 1989) foi o primeiro a decompor a variância genotípica de uma população alógama em três componentes: variância genética aditiva; variância genética de dominância; variância genética epistática.

De acordo com Falconer (1993) a importância e definição dos componentes de variância genética são:

- Variância genética aditiva: Considerando um número de gametas (célula sexual), todos carregando um mesmo gene B , que se unem aleatoriamente com gametas da população, então, a média dos genótipos produzidos se desvia da média da população pela quantidade do efeito médio do gene B . A variância aditiva é relativa ao efeito médio dos alelos, sendo a de maior importância, pois é a maior causa de semelhança entre parentes, já que os pais passam seus alelos, e não os seus genótipos, para sua progênie, ou seja, é o efeito médio dos alelos parentais que determina a média do valor genotípico de sua progênie;
- Variância genética de dominância: Considerando-se apenas um locus, a diferença entre o valor genotípico e o valor aditivo, de um genótipo particular, é conhecido como desvio de dominância, o qual representa o efeito de colocarem-se alelos em pares para formar genótipos, não levando em consideração os efeitos desses alelos separadamente. Do ponto de vista estatístico os desvios de dominância são interações entre os alelos;
- Variância genética epistática: Quando o genótipo se refere a mais de um locus, o valor genotípico pode conter um desvio extra, devido à não aditividade das combinações. Se o desvio de interação for nulo, diz-se que os genes em questão agem aditivamente entre

¹FISHER, R.A., The correlation between relatives on the supposition of Mendelian inheritance, Edinburgh, 52, 399-433, 1918.

loci. Para esse trabalho não há o interesse em estudar a variância genética epistática, sendo assim ela deve ser considerada como nula desse ponto em diante.

2.3 Coeficiente de Herdabilidade

Segundo Falconer (1993) a herdabilidade de um caráter corresponde à importância relativa das variações genotípicas na determinação de valores fenotípicos. No entanto, um caráter pode ser hereditário no sentido de ser determinado pelo genótipo, ou no sentido de ser transmitido de pai para filho, não necessariamente, esses dois sentidos são condizentes. A herdabilidade (h^2) é, então, dividida em duas, no sentido amplo e no sentido restrito, de maneira a satisfazer a ambos os sentidos possíveis para hereditariedade. Tem-se, então:

- Herdabilidade no sentido amplo: expressa o quanto das variações fenotípicas (V_f) da população é determinada pelas variações genotípicas (V_g), e pode ser estimada por

$$h^2 = \frac{V_g}{V_f}; \quad (1)$$

- Herdabilidade no sentido restrito: expressa o quanto das variações fenotípicas é determinado pela variância dos efeitos dos alelos (V_a), e pode ser estimada por

$$h^2 = \frac{V_a}{V_f}. \quad (2)$$

A herdabilidade no sentido restrito determina o grau de semelhança entre parentes, sendo, portanto, de grande importância no melhoramento genético; já a herdabilidade no sentido amplo possui um maior interesse teórico do que prático. Logo, a herdabilidade que tem valor maior para esse estudo é a no sentido restrito, sendo assim, ao deparar-se com o termo herdabilidade no texto escrito de agora em diante deve-se subentender herdabilidade no sentido restrito.

2.4 Grau Médio de Dominância

Um outro parâmetro genético de interesse é o grau médio de dominância, o qual pode ser encontrado por meio da seguinte função das variâncias genéticas:

$$GMD = \sqrt{4V_d/2V_a}, \quad (3)$$

em que GMD representa o grau médio de dominância e V_d e V_a representam as variâncias genéticas de dominância e aditiva, respectivamente.

Segundo Kearsey e Pooni (1998) o grau médio de dominância pode ser visto como uma média ponderada entre as variâncias de dominância e aditiva de todo os loci estudados. Se o efeito de dominância fosse o mesmo entre os alelos de cada locus o grau médio de dominância seria o grau de dominância valido para cada par de alelos estudados, porém o grau de dominância costuma ser distinto para diferentes loci, sendo assim o grau médio de dominância determina a importância do efeito de dominância geral em relação aos desvios aditivos dos genes.

O grau médio de dominância pode ser interpretado como parcial ($0 < GMD < 1$), completo ($GMD = 1$) ou de sobredominância ($GMD > 1$).

2.5 Delineamento Genético

O delineamento genético utilizado por Silva (2002) é conhecido como Delineamento III. Proposto por Comstock e Robinson (1952) o Delineamento III visa a estimação do grau médio de dominância para loci controladores de caracteres quantitativos (QTL). A princípio, esse delineamento propõe a utilização de indivíduos F_2 provenientes do cruzamento de duas linhagens endogâmicas divergentes quanto aos caracteres quantitativos de interesse. Tais indivíduos devem ser, então, retrocruzados com ambas as linhagens parentais, fornecendo as progênies de retrocruzamento que deverão ser observadas.

Esse delineamento possui a vantagem de permitir a estimação dos componentes de variância de dominância e aditiva de maneira independente e com precisão igual (KEARSEY; POONI, 1998). A fim de entender melhor como funciona a estimação pode-se estudar o caso para um único locus com dois alelos; o caso geral é um pouco mais complexo, mas segue a mesma linha de raciocínio.

Considere duas linhagens endogâmicas divergentes para algum caráter quantitativo de interesse L_1 e L_2 , para o caso de um único locus com dois alelos. O esquema de cruzamentos é dado pela Figura 1:

Figura 1 - Esquema de cruzamentos para a obtenção da população F_2 , em que as letras L e F significam linhagem e progênie, respectivamente; as letras B e b representam o genótipo das plantas; os valores entre parênteses representam as frequências do genótipo, quando mais de um é possível; $\times$ representa cruzamento entre as linhagens e $\otimes$ representa a autofecundação das progênies

Após obtida a população F_2 , deve-se retrocruzá-la com as duas linhagens parentais a fim de obter as plantas que deverão ser, então, observadas. Tal esquema é dado pela Figura 2:

Figura 2 - Esquema do retrocruzamento entre a população F_2 e as linhagens parentais, em que RCP representa progênie obtida por meio de retrocruzamento para os três casos possíveis; as letras B e b representam o genótipo das plantas; os valores entre parênteses representam as frequências do genótipo, quando mais de um é possível

Podem ocorrer três genótipos na população obtida por meio do retrocruzamento (RCP), BB , Bb e bb , que podem ser representados esquematicamente pela Figura 3:

Figura 3 - Esquema dos genótipos e valores genotípicos de uma população considerando um só locus. Em que BB , Bb e bb representam os possíveis genótipos, μ representa o ponto médio entre os genótipos homozigóticos, a mede o afastamento de cada genótipo homozigótico em relação à média e d mede o afastamento do heterozigoto em relação à média

A fim de ter uma idéia da variabilidade genética das progênes obtidas por meio do retrocruzamento, são dadas as Tabelas 1 e 2 com os genótipos possíveis, suas frequências e os valores genotípicos respectivos.

Tabela 1 - Genótipos, frequências relativas em que esses genótipos ocorrem e seus valores genotípicos considerando a população de retrocruzamento 1 (RCP₁)

Genótipo	Frequência	Valor Genotípico
<i>BB</i>	$\frac{1}{2}$	<i>a</i>
<i>Bb</i>	$\frac{1}{2}$	<i>d</i>

Tabela 2 - Genótipos, frequências relativas em que esses genótipos ocorrem e seus valores genotípicos considerando a população de retrocruzamento 2 (RCP₂)

Genótipo	Frequência	Valor Genotípico
<i>Bb</i>	$\frac{1}{2}$	<i>d</i>
<i>bb</i>	$\frac{1}{2}$	$-a$

É possível verificar, por meio das Tabelas 1 e 2 e das Figuras 1 e 2, que as frequências dos genótipos da população RCP não difere das frequências da população F₂; porém, a maneira como as frequências estão dispostas permite um estudo das variâncias aditiva e de dominância de maneira independente quando montada a análise da variância para o caráter quantitativo de interesse. Pois qualquer que seja o delineamento estatístico, utilizado pelo pesquisador, existirão as causas de variação devidas à progênie F₂, às linhagens L₁ e L₂ e à interação entre progênie e linhagens. Como ambas as linhagens não representam uma população aleatória, seu efeito é considerado fixo. Sendo assim, não há um componente de variância a ele relacionado; já o efeito de progênie é aleatório, pois representa uma população. Logo, o efeito da interação também é aleatório. A variância genética aditiva (V_a) pode ser encontrada por meio de uma função do componente de variância relativo à progênes (σ_p^2) e a variância genética de dominância (V_d) pode ser encontrada por meio do componente de variância relativo à interação entre progênes e linhagens (σ_{pt}^2). As funções que relacionam esses componentes, segundo Comstock e Robinson (1952) são:

$$V_a = 4\sigma_p^2 \quad (4)$$

e

$$V_d = \sigma_{pt}^2. \quad (5)$$

Existem, também, os Delineamentos I e II, propostos por Comstock e Robinson. Esses propõem, respectivamente, a utilização de indivíduos F_2 de uma maneira hierárquica, formando grupos de machos F_2 que deveriam ser cruzados com um número igual de fêmeas escolhidas aleatoriamente da população F_2 ; e de uma maneira fatorial, de modo que machos e fêmeas, da geração F_2 escolhidos aleatoriamente sejam cruzados, todos, entre si.

Diferentemente do Delineamento III, os Delineamentos I e II não fornecem estimativas das variâncias aditiva e de dominância de maneira tão direta; no Delineamento I a variância aditiva está dividida entre as variâncias relativas aos efeitos de macho e de fêmea dentro de macho, sendo que a variância de dominância também está contida na variância devida ao efeito de fêmeas dentro de macho; já no Delineamento II a variância aditiva é dada pelas variâncias relativas aos efeitos de fêmeas e de machos, sendo necessário fazer uma ponderação entre os dois valores para obter o valor da variância aditiva estimada. A variância de dominância é função apenas da variância devida à interação (COMSTOCK; ROBINSON, 1948). Também, segundo Comstock e Robinson (1952), o Delineamento III aparenta ser o mais útil, pois é o de maior poder.

2.6 Delineamento Experimental

O delineamento experimental escolhido por Silva (2002) para instalar seu experimento foi o delineamento em látice quadrado, o qual se assemelha, em alguns pontos, com os delineamentos de blocos incompletos balanceados, sendo de grande importância em experimentos com número de tratamentos elevado (GOMES, 1958) como costuma ser o caso nas fases iniciais de programas de melhoramento.

O número de tratamentos estudados por um delineamento em látice quadrado deve ser um quadrado perfeito (4, 9, 16, 25, 36, ...) esses n^2 tratamentos, com n pertencente ao conjunto dos naturais, são arranjados, na área experimental, em um quadrado $n \times n$. O método de agrupar os tratamentos em linhas e colunas, que variam de acordo com as

repetições (podendo manter a ortogonalidade entre experimentos ou não), é tal que a média de tratamentos pode ser ajustada para as diferenças entre linhas e colunas de cada látice (COCHRAN; COX, 1957). Ou seja, assim como para blocos incompletos, pode-se fazer uma análise intra-blocos ou uma análise com recuperação de informação interblocos.

2.7 Métodos de Estimação de Componentes de Variância

Existem diversos métodos para estimar componentes de variância e dentre eles o mais utilizado é o método dos momentos, ou método da análise da variância (BARBIN, 1993). O início da utilização do método da análise da variância (ANOVA) se deu com o livro de Fisher¹ (1925, Sec.40 apud SEARLE; CASELLA; McCULLOCH, 1992). Desde então, o método da ANOVA vem sendo, tradicionalmente, utilizado na estimação de componentes de variância e de coeficientes de herdabilidade relacionados ao melhoramento genético de milho, como pode ser visto em Moll; Lindsey e Robinson (1964), Arias e Souza Júnior (1998), Wolf; Peterneli e Hallauer (2000) e Silva et al. (2004).

Devido ao surgimento de novos métodos computacionais e ao aparecimento de softwares estatísticos de simples utilização, alguns métodos de estimação que eram considerados, matematicamente, inviáveis passaram a ser utilizados na estimação de componentes de variância e coeficientes de herdabilidade. Dentre esses novos métodos está o da máxima verossimilhança restrita (REML). Boca e Cantet (2004) e Wardyn; Edwards e Lamkey (2007) fazem uso do método REML em dados de milho. Pode-se perceber que os trabalhos mais novos tendem a utilizar o método da REML.

2.7.1 Método da Análise da Variância

O método da ANOVA é aquele que equaciona os quadrados médios da análise da variância com as suas respectivas esperanças matemáticas. Esse método possui diversas vantagens. Algumas delas, considerando experimentos balanceados, são: estimativas não viesadas; de variância mínima; estatísticas suficientes e completas.

Assim como vantagens, o método da ANOVA possui, também, desvantagens, sendo sua maior desvantagem o possível surgimento de estimativas negativas, já que os com-

¹FISHER, R.A. **Statistical methods for research workers**. 1.ed. Edinburgh: Oliver & Boyd, 1925.

ponentes de variância estão definidos dentro do conjunto dos Reais não negativos (SEARLE; CASELLA; McCULLOCH, 1992).

2.7.2 Método da Máxima Verossimilhança

O método da máxima verossimilhança (ML) necessita de mais informações sobre os dados do que o método da ANOVA, já que não é possível utilizar o método da ML se não for conhecida a distribuição dos dados. Esse método usa como estimativas dos valores de interesse, aqueles que maximizam a função de verossimilhança da variável aleatória em questão (SEARLE; CASELLA; McCULLOCH, 1992).

As estimativas encontradas pelo método da ML possuem tanto vantagens quanto desvantagens. Considerando experimentos balanceados, algumas das vantagens são: os estimadores encontrados pelo método ML solucionam o problema de estimativas negativas (OTSUK, 1991); são consistentes, assintoticamente normais e eficientes (FREITAS, 1991); e o procedimento de estimação é bem definido (FERNANDEZ, 1991). Porém, as estimativas calculadas pelo método da ML são tendenciosas e truncadas (OTSUK, 1991).

Uma das causas do viés encontrado nos estimadores da ML é o fato deste método não levar em consideração a perda de graus de liberdade devido a estimação dos efeitos fixos pertencentes ao modelo matemático de interesse (CUSTÓDIO, 2004), pois mesmo os modelos matemáticos aleatórios possuem um efeito fixo, aquele relativo (usualmente) à média geral das observações.

2.7.2.1 Método da Máxima Verossimilhança Restrita

O método da máxima verossimilhança restrita (REML) difere do método da ML pois leva em consideração a perda de graus de liberdade na estimação de parâmetros fixos (RAO, 1999). A idéia inicial para a estimação pelo método da REML partiu de Thompson Jr.¹ (1962, apud SEARLE; CASELLA; McCULLOCH, 1992), idéia essa de maximizar a parte da função de verossimilhança que é invariante aos efeitos fixos do modelo, estimando os componentes de variância com base em resíduos calculados após um ajuste por mínimos

¹THOMPSON JUNIOR, W.A. The problem of negative estimates of variance components **Annals of Mathematical statistics**, 33, 1, 273-289, mar. 1962.

quadrados para os efeitos fixos apenas.

Uma das características mais interessantes do método da REML, segundo Searle; Casella e McCulloch (1992) é que para determinados casos, considerando experimentos balanceados, as soluções das equações da REML são idênticas às encontradas pelo método da ANOVA, se tornando um método cada vez mais utilizado.

Em um estudo feito por Carneiro Júnior et al. (2004) são utilizados dados simulados na comparação dos métodos de estimação REML, ML e o método III de Henderson para componentes de variância de dados animais. Concluiu-se que o método REML pode ser considerado como o mais apropriado para estimar componentes de variância para caracteres de baixa herdabilidade em modelos animais. Porém, não existem estudos nessa área para dados vegetais. Dada a praticidade e tradicionalidade do método da ANOVA parece válida a tentativa de fazer a comparação entre esses dois métodos, a fim de descobrir se um é superior ao outro, considerando um modelo vegetal, mais especificamente, um modelo de melhoramento de milho.

3 MATERIAL E MÉTODOS

3.1 Material

3.1.1 Dados Reais

Os dados que serão utilizados nesse trabalho foram fornecidos pelo Departamento de Genética da Escola Superior de Agricultura "Luiz de Queiroz". Todo o processo de coleta dos dados está detalhado em Silva (2002), porém, uma parte é resumida a seguir:

3.1.1.1 Material Genético

Para a obtenção do material genético foram utilizadas duas linhagens contrastantes para diversos caracteres, principalmente para produção de grãos:

- L-08-05F, apresenta grãos duros e alaranjados (Linhagem 1 - L_1);
- L-14-04B, apresenta grãos dentados e amarelados (Linhagem 2 - L_2).

As linhagens L_1 e L_2 foram cruzadas entre si para a obtenção da geração F_1 . Por sua vez, as plantas F_1 foram autofecundadas para a obtenção da população F_2 . Foram amostradas, aleatoriamente, 250 plantas da população F_2 . Posteriormente, para o aumento do número de sementes, as plantas F_2 foram autofecundadas obtendo-se progênie $F_{2:3}$ (Figura 4). Cada progênie $F_{2:3}$ representa, em média, a planta F_2 da qual foi originada. As progênie $F_{2:3}$ foram, então, retrocruzadas com ambas as linhagens parentais da população, obtendo-se 250 progênie de retrocruzamento com cada parental (Figura 5).

$$\begin{array}{c} L_1 \\ \times \\ \\ L_2 \end{array} \longrightarrow F_1 \longrightarrow \otimes \longrightarrow F_2 \left\{ \begin{array}{l} P_1:F_2 \longrightarrow \otimes \longrightarrow P_1:F_{2:3} \\ P_2:F_2 \longrightarrow \otimes \longrightarrow P_2:F_{2:3} \\ \vdots \\ P_{250}:F_2 \longrightarrow \otimes \longrightarrow P_{250}:F_{2:3} \end{array} \right.$$

Figura 4 - Esquema para a obtenção das 250 plantas e populações utilizadas, em que as letras L, F e P representam linhagem, geração filial e planta respectivamente; $\times$ representa cruzamento entre as linhagens e $\otimes$ representa a autofecundação das progênie

$$\begin{array}{c} P_i:F_{2:3} \\ \times \\ L_1 \end{array} \left\{ \begin{array}{c} RCP_{1,1} \\ RCP_{1,2} \\ \vdots \\ RCP_{1,250} \end{array} \right., \quad \begin{array}{c} P_i:F_{2:3} \\ \times \\ L_2 \end{array} \left\{ \begin{array}{c} RCP_{2,1} \\ RCP_{2,2} \\ \vdots \\ RCP_{2,250} \end{array} \right.$$

Figura 5 - Esquema para a obtenção das 500 progênies de retrocruzamento com ambas as linhagens parentais, em que RCP representa progênie obtida através de retrocruzamento e i varia de 1 a 250

As 500 progênies de retrocruzamento são consideradas como os tratamentos que devem ser observados durante o experimento, sendo que elas são formadas a partir de dois fatores, progênies com 250 níveis e linhagens com 2 níveis.

3.1.1.2 Instalação dos Experimentos

Foram considerados 6 ambientes distintos, sendo que cada ambiente foi definido como uma combinação entre local (todas as estações experimentais pertencem à Escola Superior de Agricultura "Luiz de Queiroz- USP/Piracicaba - SP), ano agrícola e época de semeadura. Os ambientes foram:

1. Estação Experimental do Departamento de Genética (E.E.LGN) e ano 1999/2000;
2. E.E.LGN, ano 2000/2001 e primeira época de semeio;
3. E.E.LGN, ano 2000/2001 e segunda época de semeio;
4. Estação Experimental Areão e ano 2000/2001;
5. Estação Experimental Caterpillar e ano 2000/2001;
6. E.E.LGN, ano 2000/2001 e terceira época de semeio.

Em cada ambiente foram instalados 5 experimentos em látice quadrado 10×10 , com duas repetições cada. Foi realizado um sorteio para a ordenação desses experimentos sendo que a ortogonalidade entre repetições foi respeitada. Em cada um dos 5 experimentos,

repetidos 2 vezes, foram estudados 100 tratamentos, sendo que esses 100 tratamentos foram obtidos do retrocruzamento das mesmas 50 progênies F_2 com as linhagens L_1 e L_2 (Figura 6).

$$\text{Exp1} \left\{ \begin{array}{cc} \text{RCP}_{1,1} & \text{RCP}_{2,1} \\ \text{RCP}_{1,2} & \text{RCP}_{2,2} \\ \vdots & \vdots \\ \text{RCP}_{1,50} & \text{RCP}_{2,50} \end{array} \right\}, \quad \dots, \quad \text{Exp5} \left\{ \begin{array}{cc} \text{RCP}_{1,201} & \text{RCP}_{2,201} \\ \text{RCP}_{1,202} & \text{RCP}_{2,202} \\ \vdots & \vdots \\ \text{RCP}_{1,250} & \text{RCP}_{2,250} \end{array} \right\}$$

Figura 6 - Esquema dos 5 experimentos montados com 100 progênies cada, em que Exp1 e Exp5 representam os experimentos 1 e 5, respectivamente, e $\text{RCP}_{j,i}$ representam as progênies resultantes do retrocruzamento da progênie $P_i:F_{2:3}$, i variando de 1 a 250, com a linhagem j , j variando de 1 a 2

As parcelas experimentais foram formadas por uma linha de $4\text{m} \times 0,8\text{m}$. Foram semeadas 40 sementes (2 a 2), espaçadas 0,2 metros entre si em cada parcela. Após um período de 25 a 30 dias foi feito um desbaste, de maneira a manter 20 plantas por parcela. O estande (número de plantas) ideal foi de 62.500 por hectare. Para esse trabalho são considerados os dados relativos aos caracteres: produção de grãos (gramas por parcela); altura média entre cinco plantas competitivas dentro de uma mesma parcela, medida do nível do solo ao nó da inserção da folha bandeira (centímetros por planta); e número de folhas situadas entre a primeira espiga (ou espiga superior) e o pendão, o valor considerado foi a média do número de folhas para cinco plantas competitivas dentro de cada parcela.

3.2 Métodos

3.2.1 Análises Estatísticas para Delineamentos em Látice Quadrado

Um experimento montado em látice pode ser analisado de quatro maneiras diferentes (SILVA, 1997):

- Análise como blocos casualizados completos;
- Análise intrablocos, com tratamentos ajustados e blocos dentro de repetições não-ajustados;

- Análise intrablocos, com tratamentos não-ajustados e blocos dentro de repetições ajustados;
- Análise com recuperação de informação interblocos.

Silva (1997) propõe a discussão de qual dos 4 métodos de análise é o melhor quando o interesse é estimar componentes de variâncias genéticos. Em seu trabalho ele reporta a existência de algumas pequenas diferenças entre os métodos, porém não fornece uma conclusão de qual o mais apropriado, dando ênfase à necessidade de trabalhos de comparação entre as 4 diferentes análises utilizando dados simulados.

Um experimento em látice pode ser estudado de maneira individual, caso as repetições concentrem-se todas em um mesmo ambiente; ou conjunta, caso exista mais de um ambiente envolvido no experimento. Usualmente, a análise de maior interesse para a genética, é a análise conjunta, já que é de interesse para os melhoristas, estudar a interação entre ambiente e genótipos Ramalho (1977). No presente estudo o interesse é na análise conjunta, já que existem 6 ambientes envolvidos nas análises.

Cecon (1992) estudou três maneiras diferentes de analisar, conjuntamente, um experimento de melhoramento de milho em látice quadrado, sendo elas: análise como blocos casualizados, análise do látice com tratamentos não ajustados e erro intrabloco e análise como blocos casualizados utilizando as médias dos tratamentos ajustados da análise com recuperação da informação interblocos dos látices individuais (por local) . Chegou a conclusão de que, quando o objetivo é estimar parâmetros genéticos a análise mais indicada é a análise do látice como um látice, independentemente de sua eficiência, e não como blocos casualizados.

Já Regazzi et al. (1999) estudaram, além das análises discutidas por Cecon (1992), a análise conjunta do látice intrablocos com tratamentos ajustados e blocos dentro de repetições não ajustados; concluíram que existe uma concordância entre as diferentes análises nas estimativas do coeficiente de herdabilidade e na classificação dos materiais avaliados, embora existam algumas diferenças entre as estimativas dos componentes de variância.

Como o interesse principal desse trabalho é a comparação entre métodos de estimação para componentes de variância e parâmetros genéticos a análise selecionada para estudar os dados, reais e simulados, foi a análise em blocos casualizados, na qual cada látice é considerado como um bloco casualizado completo.

3.2.1.1 Modelo Matemático

Os cinco experimentos foram estudados conjuntamente, ou seja, foi feita a análise individual para cada experimento separadamente, depois foi verificada a homogeneidade de variâncias entre os experimentos. Após verificada a homogeneidade, pode-se pensar em estudar os experimentos conjuntamente, a fim de obter informações mais abrangentes sobre os efeitos estudados. A análise conjunta em grupos de experimentos é necessária pois as 50 progênies estudadas em cada experimento são diferentes entre si, logo elas devem ser comparadas dentro de cada experimento, e não entre experimentos.

O modelo matemático considerado é misto, pois possui efeitos fixo e aleatórios, além dos efeitos relativos à média geral e ao erro experimental (BARBIN, 1993). Os efeitos fixos são relativos à média geral e linhagens; todos os outros efeitos são considerados como aleatórios.

No modelo são indicados os efeitos, fixo ou aleatório, de cada fator, assim como os componentes de variância de cada efeito aleatório, já que segundo Barbin (1993) os efeitos aleatórios de um modelo matemático podem, por hipótese, seguir uma distribuição normal de média zero com variância dada pelo componente de variância associado ao efeito em questão.

O modelo para a análise conjunta em blocos casualizados é dado por:

$$y_{ijkln} = \mu + a_n + e_l + b_{k(l)} + p_{i(l)} + t_{j(l)} + ae_{ln} + pt_{ij(l)} + pa_{in(l)} + ta_{jn(l)} + pta_{ijn(l)} + \varepsilon_{ijkln}, \quad (6)$$

em que:

- y_{ijkln} representa a observação referente à progênie i retrocruzada com a linhagem genitora j avaliada no bloco k no experimento l e ambiente n , com $i=1, \dots, 50$, $j=1, 2$, $k=1, 2$, $l=1, \dots, 5$ e $n=1, \dots, 6$;
- μ representa a média geral das observações, em que μ é de efeito fixo, com $E(\mu) = \mu$, $E(\mu^2) = \mu^2$;
- a_n representa o efeito do ambiente n , em que a_n é de efeito aleatório, e portanto $a_n \sim NIID(0, \sigma_a^2)$, em que $NIID$ é a sigla para Normal Independentemente e Identicamente Distribuído;

- e_l representa o efeito do experimento l , em que e_l é de efeito aleatório, e portanto $e_l \sim NIID(0, \sigma_e^2)$;
- $b_{k(l)}$ representa o efeito do bloco k dentro do experimento l , em que $b_{k(l)}$ é de efeito aleatório, e portanto $b_{k(l)} \sim NIID(0, \sigma_b^2)$;
- $p_{i(l)}$ representa o efeito da progênie i dentro do experimento l , em que $p_{i(l)}$ é de efeito aleatório, e portanto $p_{i(l)} \sim NIID(0, \sigma_p^2)$;
- $t_{j(l)}$ representa o efeito da linhagem genitora j dentro do experimento l , em que $t_{j(l)}$ é de efeito fixo, com $E(t_{j(l)}) = t_{j(l)}$, $E(t_{j(l)}^2) = t_{j(l)}^2$ e $\sum_j t_{j(l)} = 0$;
- ae_{ln} representa o efeito da interação do ambiente n com o experimento l , em que ae_{ln} é de efeito aleatório, e portanto $ae_{ln} \sim NIID(0, \sigma_{ae}^2)$;
- $pt_{ij(l)}$ representa o efeito da interação da progênie i com a linhagem genitora j dentro do experimento l , em que $pt_{ij(l)}$ é de efeito aleatório, e portanto $pt_{ij(l)} \sim NIID(0, \sigma_{pt}^2)$;
- $pa_{in(l)}$ representa o efeito da interação da progênie i com o ambiente l dentro do experimento l , em que $pa_{in(l)}$ é de efeito aleatório, e portanto $pa_{in(l)} \sim NIID(0, \sigma_{pa}^2)$;
- $ta_{jn(l)}$ representa o efeito da interação da linhagem genitora j com o ambiente l dentro do experimento l , em que $ta_{jn(l)}$ é de efeito aleatório, e portanto $ta_{jn(l)} \sim NIID(0, \sigma_{ta}^2)$;
- $pta_{ijn(l)}$ representa o efeito da interação da progênie i , com a linhagem genitora j e com o ambiente n dentro do experimento l , em que $pta_{ijn(l)}$ é de efeito aleatório, e portanto $pta_{ijn(l)} \sim NIID(0, \sigma_{pta}^2)$;
- ε_{ijkl} representa o erro aleatório associado à observação y_{ijkl} , em que ε_{ijkl} é de efeito aleatório, e portanto $\varepsilon_{ijkl} \sim NIID(0, \sigma^2)$.

A restrição $\sum_j t_{j(l)} = 0$, imposta para o efeito de linhagem dentro de experimento (único efeito fixo que não a média), é usualmente utilizada na obtenção dos estimadores de componentes da variância pelo método da ANOVA, e não implica na mudança da hipótese nula H_0 , a qual considera que os tratamentos não diferem entre si (BARBIN, 1993).

Existem, porém, outras restrições que devem ser consideradas, essas são:

$$\sum_j pt_{ij(l)} = 0; \quad (7)$$

$$\sum_j ta_{jn(l)} = 0; \quad (8)$$

$$\sum_j pta_{ijn(l)} = 0. \quad (9)$$

Ao considerar essas três condições extras deve-se redefinir as seguintes esperanças matemáticas:

$$E(pt_{ij(l)}^2) = \frac{J-1}{J} \sigma_{pt}^2; \quad (10)$$

$$E(ta_{jn(l)}^2) = \frac{J-1}{J} \sigma_{ta}^2; \quad (11)$$

$$E(pta_{ijn(l)}^2) = \frac{J-1}{J} \sigma_{pta}^2. \quad (12)$$

Segundo Barbin (1993) as restrições dadas pelas eq. (7), (8) e (9) são optativas, ficando a cargo do pesquisador a escolha por utilizá-las ou não. A escolha pela utilização das restrições impostas às interações foi feita com base no delineamento genético utilizado, pois os idealizadores do Delineamento III, Comstock e Robinson (1952), consideram tais restrições na hora de estimar os componentes de variância do modelo.

A não utilização das restrições (eq. (7), (8) e (9)) resultaria em uma incongruência com a maneira que o material genético observado foi obtido, como pode ser visto no item delineamento genético (item 3.5), dentro de revisão de literatura. O efeito relativo à interação entre progênies e linhagens é relativo apenas à variância de dominância e o efeito de progênies é relativo apenas à variância aditiva, não dependendo da variância de dominância, ou seja, não deve estar diretamente relacionado ao efeito da interação entre progênies e linhagens, e a utilização das restrições nos efeitos de interação garante essa independência.

3.2.1.2 Estimação dos Parâmetros Aleatórios

3.2.1.2.1 Método da Análise da Variância

Para obter os estimadores dos componentes de variância pelo método da ANOVA é necessário definir as equações das somas de quadrados e quadrados médios da

ANOVA, a fim de calcular as esperanças dos quadrados médio. A Tabela 3 fornece as causas de variação e seus respectivos graus de liberdade para o modelo (6).

Tabela 3- ANOVA para a análise conjunta dos dados segundo o delineamento fatorial implantado em blocos casualizados, modelo (6), em que CV representa a causa de variação e GL os graus de liberdade, sendo que d. significa dentro de

CV	GL
Ambientes (A)	$N - 1$
Experimentos (E)	$L - 1$
A×E	$(N - 1)(L - 1)$
Blocos d. A d. E	$(K - 1)LN$
Progenies (P) d. E	$(I - 1)L$
Linhagens (T) d. E	$(J - 1)L$
T×P d. E	$(I - 1)(J - 1)L$
P×A d. E	$(I - 1)(N - 1)L$
T×A d. E	$(J - 1)(N - 1)L$
P×T×A d. E	$(I - 1)(J - 1)(N - 1)L$
Resíduos	$(IJLN - LN)(K - 1)$
Total	$IJKLN - 1$

Em que: $N = 6$, $L = 5$, $K = 2$, $I = 50$, $J = 2$.

As somas de quadrados das causas de variação são dadas por:

$$SQ_{Total} = \sum_{i,j,k,l,n} y_{ijkln}^2 - C \quad (13)$$

$$C = \frac{(\sum_{i,j,k,l,n} y_{ijkln})^2}{IJKLN} \quad (14)$$

$$SQA = \frac{1}{IJKL} \sum_n (\sum_{i,j,k,l} y_{ijkln})^2 - C \quad (15)$$

$$SQE = \frac{1}{IJKN} \sum_l (\sum_{i,j,k,n} y_{ijkln})^2 - C \quad (16)$$

$$\text{SQAE} = \frac{1}{IJK} \sum_{l,n} \left(\sum_{i,j,k} y_{ijkln} \right)^2 - C - \text{SQA} - \text{SQE} \quad (17)$$

$$\text{SQB} = \frac{1}{IJ} \sum_{l,n} \left[\sum_k \left(\sum_{i,j} y_{ijkln} \right)^2 \right] - \frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln} \right)^2 \quad (18)$$

$$\text{SQP} = \frac{1}{IJN} \sum_l \left[\sum_i \left(\sum_{j,k,n} y_{ijkln} \right)^2 \right] - \frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln} \right)^2 \quad (19)$$

$$\text{SQT} = \frac{1}{IKN} \sum_l \left[\sum_j \left(\sum_{i,k,n} y_{ijkln} \right)^2 \right] - \frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln} \right)^2 \quad (20)$$

$$\text{SQPT} = \frac{1}{KN} \sum_l \left[\sum_{i,j} \left(\sum_{k,n} y_{ijkln} \right)^2 \right] - \frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln} \right)^2 - \text{SQP} - \text{SQT} \quad (21)$$

$$\text{SQPA} = \frac{1}{JK} \sum_l \left[\sum_{i,n} \left(\sum_{j,k} y_{ijkln} \right)^2 \right] - \frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln} \right)^2 - \text{SQP} - \text{SQA} - \text{SQAE} \quad (22)$$

$$\text{SQTA} = \frac{1}{IK} \sum_l \left[\sum_{j,n} \left(\sum_{i,k} y_{ijkln} \right)^2 \right] - \frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln} \right)^2 - \text{SQT} - \text{SQA} - \text{SQAE} \quad (23)$$

$$\begin{aligned} \text{SQPTA} = & \frac{1}{K} \sum_l \left[\sum_{i,j,n} \left(\sum_k y_{ijkln} \right)^2 \right] - \frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln} \right)^2 - \text{SQP} - \text{SQT} + \\ & - \text{SQA} - \text{SQPT} - \text{SQPA} - \text{SQTA} - \text{SQAE} \end{aligned} \quad (24)$$

$$\begin{aligned} \text{SQRes} = & \text{SQTotal} - \text{SQA} - \text{SQE} - \text{SQAE} - \text{SQB} - \text{SQP} - \text{SQT} - \text{SQPT} + \\ & - \text{SQPA} - \text{SQTA} - \text{SQPTA} \end{aligned} \quad (25)$$

O passo seguinte na estimação pelo método da ANOVA é aplicar a esperança matemática a cada uma das somas de quadrados das causas de variação do modelo; porém tal aplicação não é tão simples, a fim de minimizar a possibilidade de erros nas contas foi aplicada a esperança matemática em cada um dos conjuntos de somatórios separadamente. Com fim ilustrativo é detalhada a aplicação da esperança matemática ao primeiro conjunto de somatórios, sendo que para os demais é dado apenas o resultado.

$$\begin{aligned}
E\left(\sum_{i,j,k,l,n} y_{ijkln}^2\right) &= \sum_{i,j,k,l,n} [E(\mu + a_n + e_l + b_{k(l)} + p_{i(l)} + t_{j(l)} + ae_{ln} + pt_{ij(l)} + pa_{in(l)} + \\
&\quad + ta_{jn(l)} + pta_{ijn(l)} + \varepsilon_{ijkln})^2] \\
&= \sum_{i,j,k,l,n} [E(\mu^2 + a_n^2 + e_l^2 + b_{k(l)}^2 + p_{i(l)}^2 + t_{j(l)}^2 + ae_{ln}^2 + pt_{ij(l)}^2 + pa_{in(l)}^2 + \\
&\quad + ta_{jn(l)}^2 + pta_{ijn(l)}^2 + \varepsilon_{ijkln}^2 + dp)] \\
&= IJKL(\mu^2 + \sigma_a^2 + \sigma_e^2 + \sigma_b^2 + \sigma_p^2 + t_{j(l)}^2 + \sigma_{ae}^2 + \frac{(J-1)}{J}\sigma_{pt}^2 + \sigma_{pa}^2 + \\
&\quad + \frac{(J-1)}{J}\sigma_{ta}^2 + \frac{(J-1)}{J}\sigma_{pta}^2 + \sigma^2) \\
&= IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKLN\sigma_e^2 + IJKLN\sigma_b^2 + IJKLN\sigma_p^2 + \\
&\quad + IKLN \sum_{j,l} t_{j(l)}^2 + IJKLN\sigma_{ae}^2 + IKLN(J-1)\sigma_{pt}^2 + IJKLN\sigma_{pa}^2 + \\
&\quad + IKLN(J-1)\sigma_{ta}^2 + IKLN(J-1)\sigma_{pta}^2 + IJKLN\sigma^2
\end{aligned} \tag{26}$$

em que dp representa os duplos produtos obtidos ao elevar o modelo ao quadrado. Como o modelo (6) é considerado aditivo, exigência das pressuposições da ANOVA, os efeitos são considerados como independentes entre si, o que significa que a esperança dos duplos produtos é nula.

$$\begin{aligned}
E\left[\frac{1}{IJKLN} \left(\sum_{i,j,k,l,n} y_{ijkln}\right)^2\right] &= IJKLN\mu^2 + IJKL\sigma_a^2 + IJKN\sigma_e^2 + IJ\sigma_b^2 + JKN\sigma_p^2 + \\
&\quad + IJK\sigma_{ae}^2 + JK\sigma_{pa}^2 + \sigma^2
\end{aligned} \tag{27}$$

$$\begin{aligned}
E\left[\frac{1}{IJKL} \sum_n \left(\sum_{i,j,k,l} y_{ijkln}\right)^2\right] &= IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKN\sigma_e^2 + IJN\sigma_b^2 + \\
&\quad + JKN\sigma_p^2 + IJKN\sigma_{ae}^2 + JKN\sigma_{pa}^2 + N\sigma^2
\end{aligned} \tag{28}$$

$$\begin{aligned}
E\left[\frac{1}{IJKN} \sum_l \left(\sum_{i,j,k,n} y_{ijkln}\right)^2\right] &= IJKLN\mu^2 + IJKL\sigma_a^2 + IJKLN\sigma_e^2 + IJL\sigma_b^2 + \\
&\quad + JKLN\sigma_p^2 + IJKL\sigma_{ae}^2 + JKL\sigma_{pa}^2 + L\sigma^2
\end{aligned} \tag{29}$$

$$\begin{aligned}
E\left[\frac{1}{IJK} \sum_{l,n} \left(\sum_{i,j,k} y_{ijkln}\right)^2\right] &= IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKLN\sigma_e^2 + IJLN\sigma_b^2 + \\
&\quad + JKLN\sigma_p^2 + IJKLN\sigma_{ae}^2 + JKLN\sigma_{pa}^2 + LN\sigma^2
\end{aligned} \tag{30}$$

$$\begin{aligned} \mathbb{E} \frac{1}{IJ} \sum_{l,n} [\sum_k (\sum_{i,j} y_{ijkln})^2] &= IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKLN\sigma_e^2 + IJKLN\sigma_b^2 + \\ &+ JKLN\sigma_p^2 + IJKLN\sigma_{ae}^2 + JKLN\sigma_{pa}^2 + KLN\sigma^2 \end{aligned} \quad (31)$$

$$\begin{aligned} \mathbb{E} \frac{1}{IJN} \sum_l [\sum_i (\sum_{j,k,n} y_{ijkln})^2] &= IJKLN\mu^2 + IJKL\sigma_a^2 + IJKLN\sigma_e^2 + IJL\sigma_b^2 + \\ &+ IJKLN\sigma_p^2 + IJKL\sigma_{ae}^2 + IJKL\sigma_{pa}^2 + IL\sigma^2 \end{aligned} \quad (32)$$

$$\begin{aligned} \mathbb{E} \frac{1}{IKN} \sum_l [\sum_j (\sum_{i,k,n} y_{ijkln})^2] &= IJKLN\mu^2 + IJKL\sigma_a^2 + IJKLN\sigma_e^2 + IJL\sigma_b^2 + \\ &+ JKLN\sigma_p^2 + IKN \sum_{j,l} t_{j(l)}^2 + IJKL\sigma_{ae}^2 + \\ &+ KLN(J-1)\sigma_{pt}^2 + IJKL\sigma_{pa}^2 + IKL(J-1)\sigma_{ta}^2 + \\ &+ KL(J-1)\sigma_{pta}^2 + JL\sigma^2 \end{aligned} \quad (33)$$

$$\begin{aligned} \mathbb{E} \frac{1}{KN} \sum_l [\sum_{i,j} (\sum_{k,n} y_{ijkln})^2] &= IJKLN\mu^2 + IJKL\sigma_a^2 + IJKLN\sigma_e^2 + IJL\sigma_b^2 + \\ &+ IJKLN\sigma_p^2 + IKN \sum_{j,l} t_{j(l)}^2 + IJKL\sigma_{ae}^2 + \\ &+ IKLN(J-1)\sigma_{pt}^2 + IJKL\sigma_{pa}^2 + IKL(J-1)\sigma_{ta}^2 + \\ &+ IKL(J-1)\sigma_{pta}^2 + IJL\sigma^2 \end{aligned} \quad (34)$$

$$\begin{aligned} \mathbb{E} \frac{1}{JK} \sum_l [\sum_{i,n} (\sum_{j,k} y_{ijkln})^2] &= IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKLN\sigma_e^2 + IJLN\sigma_b^2 + \\ &+ IJKLN\sigma_p^2 + IJKLN\sigma_{ae}^2 + IJKLN\sigma_{pa}^2 + ILN\sigma^2 \end{aligned} \quad (35)$$

$$\begin{aligned} \mathbb{E} \frac{1}{IK} \sum_l [\sum_{j,n} (\sum_{i,k} y_{ijkln})^2] &= IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKLN\sigma_e^2 + IJLN\sigma_b^2 + \\ &+ JKLN\sigma_p^2 + IKN \sum_{j,l} t_{j(l)}^2 + IJKLN\sigma_{ae}^2 + \\ &+ KLN(J-1)\sigma_{pt}^2 + JKLN\sigma_{pa}^2 + IKLN(J-1)\sigma_{ta}^2 + \\ &+ KLN(J-1)\sigma_{pta}^2 + JLN\sigma^2 \end{aligned} \quad (36)$$

$$\begin{aligned}
E \frac{1}{K} \sum_l \left[\sum_{i,j,n} \left(\sum_k y_{ijkln} \right)^2 \right] &= IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKLN\sigma_e^2 + IJLN\sigma_b^2 + \\
&+ IJKLN\sigma_p^2 + IKN \sum_{j,l} t_{j(l)}^2 + IJKLN\sigma_{ae}^2 + \\
&+ IKLN(J-1)\sigma_{pt}^2 + IJKLN\sigma_{pa}^2 + IKLN(J-1)\sigma_{ta}^2 + \\
&+ IKLN(J-1)\sigma_{pta}^2 + IJLN\sigma^2
\end{aligned} \tag{37}$$

Subtraindo a eq. (27) da eq. (28) obtém-se a esperança da soma de quadrados de ambientes, dividindo o resultado pelos respectivos graus de liberdade obtém-se a esperança do quadrado médio de ambientes:

$$\begin{aligned}
E[\text{QMA}] &= \frac{1}{(N-1)} [(IJKLN\mu^2 + IJKLN\sigma_a^2 + IJKN\sigma_e^2 + IJN\sigma_b^2 + \\
&+ JKN\sigma_p^2 + IJKN\sigma_{ae}^2 + JKN\sigma_{pa}^2 + N\sigma^2) + \\
&- (IJKLN\mu^2 + IJKL\sigma_a^2 + IJKN\sigma_e^2 + IJ\sigma_b^2 + JKN\sigma_p^2 + \\
&+ IJK\sigma_{ae}^2 + JK\sigma_{pa}^2 + IJKLN\sigma^2)] \\
&= \frac{(N-1)(IJKL\sigma_a^2 + IJ\sigma_b^2 + IJK\sigma_{ae}^2 + JK\sigma_{pa}^2 + \sigma^2)}{N-1} \\
&= IJKL\sigma_a^2 + IJ\sigma_b^2 + IJK\sigma_{ae}^2 + JK\sigma_{pa}^2 + \sigma^2
\end{aligned} \tag{38}$$

Seguindo o mesmo raciocínio obtém-se todas as esperanças dos quadrados médios das causas de variação da análise da variância:

$$E[\text{QME}] = IJKN\sigma_e^2 + IJ\sigma_b^2 + JKN\sigma_p^2 + IJK\sigma_{ae}^2 + IJK\sigma_{pa}^2 + \sigma^2 \tag{39}$$

$$E[\text{QMAE}] = IJ\sigma_b^2 + IJK\sigma_{ae}^2 + JK\sigma_{pa}^2 + \sigma^2 \tag{40}$$

$$E[\text{QMB}] = IJ\sigma_b^2 + \sigma^2 \tag{41}$$

$$E[\text{QMP}] = JKN\sigma_p^2 + JK\sigma_{pa}^2 + \sigma^2 \tag{42}$$

$$E[\text{QMT}] = \frac{IKN}{L(J-1)} \sum_{j,l} t_{j(l)}^2 + KN\sigma_{pt}^2 + IK\sigma_{ta}^2 + K\sigma_{pta}^2 + \sigma^2 \tag{43}$$

$$E[\text{QMPT}] = KN\sigma_{pt}^2 + K\sigma_{pta}^2 + \sigma^2 \tag{44}$$

$$E[\text{QMPA}] = JK\sigma_{pa}^2 + \sigma^2 \tag{45}$$

$$E[\text{QMTA}] = IK\sigma_{ta}^2 + K\sigma_{pta}^2 + \sigma^2 \tag{46}$$

$$E[QMPTA] = K\sigma_{pta}^2 + \sigma^2 \quad (47)$$

$$E[QMRes] = \sigma^2 \quad (48)$$

Para obter os estimadores dos componentes da variância pelo método da ANOVA basta igualar os quadrados médios às suas esperanças. Logo, os estimadores dos componentes da variância para o modelo (6) são dados por:

$$\hat{\sigma}^2 = QMRes; \quad (49)$$

$$\hat{\sigma}_{pta}^2 = \frac{QMPTA - QMRes}{K}; \quad (50)$$

$$\hat{\sigma}_{ta}^2 = \frac{QMTA - QMPTA}{IK}; \quad (51)$$

$$\hat{\sigma}_{pa}^2 = \frac{QMPA - QMRes}{JK}; \quad (52)$$

$$\hat{\sigma}_{pt}^2 = \frac{QMPT - QMPTA}{KN}; \quad (53)$$

$$\hat{\sigma}_{ae}^2 = \frac{QMAE + QMRes - QMB - QMPA}{IJK}; \quad (54)$$

$$\hat{\sigma}_p^2 = \frac{QMP - QMPA}{JKN}; \quad (55)$$

$$\hat{\sigma}_b^2 = \frac{QMB - QMRes}{IJ}; \quad (56)$$

$$\hat{\sigma}_e^2 = \frac{QME + QMPA - QMP - QMAE}{IJKN}; \quad (57)$$

$$\hat{\sigma}_a^2 = \frac{QMA - QMAE}{IJKL}; \quad (58)$$

3.2.1.2.2 Método da Máxima Verossimilhança Restrita (REML)

O método REML não pode ser calculado de maneira tão direta quanto o método da ANOVA. Sendo assim, toda a análise foi feita com o auxílio do software SAS.

Segundo Littell et al. (2006) o procedimento PROC MIXED do software SAS aplica métodos de verossimilhança a modelos mistos, independentemente de sua complexidade. Esse procedimento fornece diferentes maneiras de estimar os componentes de variância relativos aos efeitos aleatórios de um modelo misto e dentre estas, tem-se o método da máxima verossimilhança.

Ao utilizar o PROC MIXED deve-se especificar o conjunto de dados a ser utilizado e o método de estimação logo na primeira linha, utilizando os comandos DATA e METHOD, respectivamente; na Segunda linha, deve-se especificar as causas de variação consideradas na tabela da análise da variância, utilizando o comando CLASS; na terceira linha, deve-se especificar a variável resposta e os efeitos fixos do modelo, utilizando o comando MODEL e o sinal de igual entre a variável resposta e os efeitos fixos; na penúltima linha devem estar os efeitos aleatórios, especificados pelo comando RANDOM; e na última linha o comando RUN para encerrar o procedimento.

Após processar o programa montado no editor do SAS, pode-se encontrar na página de saída (OUTPUT) os valores dos componentes de variância, obtidos pelo método REML, relativos aos efeitos aleatórios do modelo, juntamente com outras informações do modelo e do processo de estimação.

3.2.1.2.3 Preditores dos Parâmetros Genéticos

Segundo Silva (2002) os preditores dos parâmetros genéticos são:

- Variância aditiva: $\hat{V}_a = 4\hat{\sigma}_p^2$;
- Variância de dominância: $\hat{V}_d = \hat{\sigma}_{pt}^2$;
- Variância genotípica: $\hat{V}_g = 4\hat{\sigma}_p^2 + \hat{\sigma}_{pt}^2$;
- Variância fenotípica de médias de progênes de meios-irmãos: $\hat{V}_f = \hat{\sigma}_p^2 + \frac{\hat{\sigma}_{pa}^2}{JN} + \frac{\hat{\sigma}^2}{JKN}$;
- Interação entre as variâncias aditiva e do ambiente: $\hat{V}_{ae} = 4\hat{\sigma}_{pa}^2$;
- Interação entre as variâncias de dominância e do ambiente: $\hat{V}_{de} = \hat{\sigma}_{pta}^2$;
- Interação entre as variâncias genética e do ambiente: $\hat{V}_{ge} = 4\hat{\sigma}_{pa}^2 + \hat{\sigma}_{pta}^2$;
- Coeficiente de herdabilidade para médias de progênes de meios-irmãos:

$$\hat{h}^2 = \frac{\hat{\sigma}_p^2}{(\hat{\sigma}_p^2 + \frac{\hat{\sigma}_{pa}^2}{JN} + \frac{\hat{\sigma}^2}{JKN})}$$
;
- Grau médio de dominância: $\hat{GMD} = \sqrt{\frac{4(\hat{\sigma}_{pt}^2)}{2(4\hat{\sigma}_p^2)}}$.

3.2.1.3 Estimação dos Efeitos Fixos

Considere o modelo (6) na sua forma matricial:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (59)$$

em que:

- $\mathbf{y}$ é o vetor de dados coletados, portanto conhecido;
- $\mathbf{X}$ é a matriz de delineamento, de dimensão 6000×11 , que considera apenas os efeitos fixos do modelo;
- $\boldsymbol{\beta}$ é o vetor de efeitos fixos de dimensão 11×1 ;
- $\mathbf{Z}$ é a matriz de delineamento, de dimensão 6000×5411 , que considera apenas os efeitos aleatórios do modelo;
- $\boldsymbol{\theta}$ é o vetor de efeitos aleatórios de dimensão 5411×1 ;
- $\boldsymbol{\varepsilon}$ é o vetor dos erros aleatórios, associados às observações, de dimensão 6000×1 .

Após obtidos os componentes de variância dos efeitos aleatórios, incluindo o erro experimental, relativos ao modelo (6) é possível obter estimativas dos vetores $\boldsymbol{\theta}$ ($\hat{\boldsymbol{\theta}}$) e $\boldsymbol{\varepsilon}$ ($\hat{\boldsymbol{\varepsilon}}$), por meio do comando `rnorm` do software R. Ou seja, a única incógnita do modelo (59) passa a ser o vetor $\boldsymbol{\beta}$, já que as matrizes $\mathbf{X}$ e $\mathbf{Z}$ podem ser obtidas a partir das posições, tanto das observações no vetor $\mathbf{y}$, quanto dos efeitos nos vetores $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$.

Reorganizando o modelo (59) tem-se:

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{y} - \mathbf{Z}\boldsymbol{\theta} - \boldsymbol{\varepsilon} \quad (60)$$

$$\mathbf{X}^t \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^t (\mathbf{y} - \mathbf{Z}\boldsymbol{\theta} - \boldsymbol{\varepsilon}) \quad (61)$$

em que $\mathbf{X}^t$ representa a matriz $\mathbf{X}$ transposta. A matriz $\mathbf{X}^t \mathbf{X}$ é dada por:

$$\mathbf{X}^t\mathbf{X} = \begin{bmatrix} IJKLN & IJN & IJN & IJN & IJN & IJN & IJN & IJN & IJN & IJN & IJN \\ IJN & IJN & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ IJN & 0 & IJN & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ IJN & 0 & 0 & IJN & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ IJN & 0 & 0 & 0 & IJN & 0 & 0 & 0 & 0 & 0 & 0 \\ IJN & 0 & 0 & 0 & 0 & IJN & 0 & 0 & 0 & 0 & 0 \\ IJN & 0 & 0 & 0 & 0 & 0 & IJN & 0 & 0 & 0 & 0 \\ IJN & 0 & 0 & 0 & 0 & 0 & 0 & IJN & 0 & 0 & 0 \\ IJN & 0 & 0 & 0 & 0 & 0 & 0 & 0 & IJN & 0 & 0 \\ IJN & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & IJN & 0 \\ IJN & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & IJN \end{bmatrix} \quad (62)$$

Para obter o estimador de mínimos quadrados de β seria necessário multiplicar ambos os lados da eq. (61) pela inversa de $\mathbf{X}^t\mathbf{X}$, porém essa é singular, logo, não possui inversa. É possível, entretanto, utilizar a restrição $\sum_j t_{j(l)} = 0$, definida para o modelo (6), na obtenção de uma matriz não singular que possa ser utilizada na obtenção das estimativas de mínimos quadrados (RENCHE, 2000).

A fim de inserir a restrição na matriz $\mathbf{X}$ basta adicionar, a ela, 5 linhas, obtendo-se uma nova matriz $\mathbf{X}^*$ de delineamento. Cada uma das 5 linhas descrevendo a restrição de que a soma das linhagens 1 e 2 deve ser nula para cada um dos 5 experimentos, para tal basta a linha conter zeros para todos os efeitos menos para as linhagens dentro do experimento ao qual a linha se relaciona. Essas duas colunas (na linha) devem receber o valor 1.

Após adicionar a restrição à matriz $\mathbf{X}$, formando a nova matriz $\mathbf{X}^*$ de dimensões 6005×11 , deve se adicionar 5 linhas aos vetores $\mathbf{y}$ e ϵ e à matriz $\mathbf{Z}$, e essas cinco linhas devem ser todas de zeros. Serão formados então, dois novos vetores $\mathbf{y}^*$ (6005×1) e ϵ^* (6005×1), e uma nova matriz $\mathbf{Z}^*$ (6005×5411).

Substituindo as novas matrizes e vetores na eq. 61, tem-se:

$$(\mathbf{X}^*)^t \mathbf{X}^* \beta = (\mathbf{X}^*)^t (\mathbf{y}^* - \mathbf{Z}^* \theta - \epsilon^*) \quad (63)$$

A nova matriz $(\mathbf{X}^*)^t \mathbf{X}^*$ é não singular, logo, possui inversa. Sendo assim, o estimador de mínimos quadrados de β é dado por:

$$\hat{\beta} = ((\mathbf{X}^*)^t \mathbf{X}^*)^{-1} (\mathbf{X}^*)^t (\mathbf{y}^* - \mathbf{Z}^* \hat{\theta} - \hat{\epsilon}^*) \quad (64)$$

Para a obtenção das estimativas dos efeitos fixos foram simulados, por meio do comando `rnorm` do software R, os vetores $\hat{\theta}$ e $\hat{\epsilon}^*$, estes foram, então, inseridos na eq. (64). Porém, esse processo não foi único. A fim de obter estimativas confiáveis para os efeitos fixos foram simulados 1000 vetores $\hat{\theta}$ e $\hat{\epsilon}^*$, respectivamente, arranjados em matrizes de maneira que cada coluna representasse um vetor. Analogamente, o vetor $\mathbf{y}^*$ foi repetido 1000 vezes. A eq. (64) fica, então:

$$\hat{\mathbf{B}} = ((\mathbf{X}^*)^t \mathbf{X}^*)^{-1} (\mathbf{X}^*)^t (\mathbf{Y} - \mathbf{Z}^* \hat{\mathbf{T}} - \hat{\mathbf{E}}) \quad (65)$$

em que:

- $\hat{\mathbf{B}}$, de dimensão 11×1000 , é a matriz de estimativas;
- $((\mathbf{X}^*)^t \mathbf{X}^*)^{-1}$, de dimensão 11×11 , é a inversa da matriz $(\mathbf{X}^*)^t \mathbf{X}^*$;
- $(\mathbf{X}^*)^t$, de dimensão 11×6005 , é a transposta da matriz $\mathbf{X}$;
- $\mathbf{Y}$, de dimensão 6005×1000 , é a matriz de dados, aumentada de zeros, cujas colunas são todas iguais ao vetor $\mathbf{y}^*$;
- $\mathbf{Z}^*$, de dimensão 6005×5411 , mantém o mesmo significado, porém aumentada de zeros;
- $\hat{\mathbf{T}}$, de dimensão 5411×1000 , é a matriz de estimativas dos efeitos aleatórios, cujas colunas representam vetores $\hat{\theta}$ de estimativas simulados independentemente;
- $\hat{\mathbf{E}}$, de dimensão 6005×1000 , é a matriz de estimativas dos erros experimentais, aumentada de zeros, cujas colunas representam vetores $\hat{\epsilon}^*$ de estimativas simulados independentemente;

É importante ressaltar que a matriz de efeitos aleatórios $\hat{\mathbf{T}}$ foi montada de maneira a respeitar as restrições feitas aos efeitos aleatórios relativos às interações entre:

linhagens e progênes ($\sum_j pt_{ij(l)} = 0$); linhagens e ambientes ($\sum_j ta_{jn(l)} = 0$); e linhagens, progênes e ambientes ($\sum_j pta_{ijn(l)} = 0$).

Foi feito um estudo descritivo das 1000 estimativas encontradas pra cada um dos efeitos fixos, a fim de verificar qual a melhor estatística a ser utilizada pra representar os efeitos fixos do modelo. Foi observado que os dados se comportavam de maneira simétrica, sendo assim, as estimativas finais dos efeitos fixos do modelo, aquelas que serão utilizadas na simulação dos dados, são dadas pela média das 1000 estimativas encontradas pelo método descrito, ou seja, pela média dos valores de cada linha da matriz $\hat{\mathbf{B}}$.

3.3 Dados Simulados

A simulação de dados é uma ferramenta muito útil para comparação de métodos de estimação de parâmetros estatísticos, pois para simular, por exemplo, uma amostra de uma distribuição normal é necessário informar ao programa qual o valor da média e da variância da distribuição. Sendo assim, na hora de estimar esses dois parâmetros têm-se os valores utilizados como uma referência dos valores que se desejam alcançar.

Os dados que serão simulados para este trabalho seguirão o mesmo molde dos dados reais já mencionados, modelo (6), de maneira que possibilite a utilização dos dados simulados como uma base para a comparação dos dois métodos de estimação de interesse.

3.3.0.4 Simulação dos Dados

A simulação dos dados será feita com o auxílio do comando `rnorm` do software R. Esse comando permite a obtenção de uma amostra aleatória, de tamanho definido, pertencente a uma distribuição normal de média e desvio padrão conhecidos.

A partir dos dados reais foram obtidas estimativas dos efeitos fixos, reunidas no vetor $\hat{\boldsymbol{\theta}}$, e foram também obtidas as estimativas dos componentes de variância relativos aos efeitos aleatórios do modelo (6) ($\hat{\sigma}^2$, $\hat{\sigma}_{pt}^2$, $\hat{\sigma}_p^2$, $\hat{\sigma}_b^2$ e $\hat{\sigma}_e^2$).

O primeiro passo para obter conjuntos de dados simulados no software R é montar o modelo matemático, ao qual os dados devem respeitar, na sua forma matricial. Tal modelo é igual ao utilizado para estimar os efeitos fixos, e é dado pela eq. (59). As matrizes

de delineamento são as mesmas, e fornecem informação sobre como devem ser montados os vetores de efeitos fixos e aleatórios.

O vetor de efeito fixo deve ser montado da seguinte maneira:

$$\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \tilde{\mu} \\ \tilde{t}_{1(1)} \\ \tilde{t}_{2(1)} \\ \tilde{t}_{1(2)} \\ \tilde{t}_{2(2)} \\ \tilde{t}_{1(3)} \\ \tilde{t}_{2(3)} \\ \tilde{t}_{1(4)} \\ \tilde{t}_{2(4)} \\ \tilde{t}_{1(5)} \\ \tilde{t}_{2(5)} \end{bmatrix} \quad (66)$$

O vetor de efeito aleatório deve ser montado da seguinte maneira:

$$\tilde{\theta} = \begin{bmatrix} \tilde{\mathbf{a}} \\ \dots \\ \tilde{\mathbf{e}} \\ \dots \\ \mathbf{\tilde{a}e} \\ \dots \\ \tilde{\mathbf{b}} \\ \dots \\ \tilde{\mathbf{p}} \\ \dots \\ \mathbf{\tilde{p}t} \\ \dots \\ \tilde{\mathbf{t}a} \\ \dots \\ \mathbf{\tilde{p}a} \\ \dots \\ \mathbf{\tilde{p}ta} \end{bmatrix} \quad (67)$$

em que:

- $\tilde{\mathbf{a}}$ é o vetor dos efeitos aleatórios de ambientes, de dimensão 6×1 ;
- $\tilde{\mathbf{e}}$ é o vetor dos efeitos aleatórios de experimentos, de dimensão 5×1 ;
- $\mathbf{\tilde{a}e}$ é o vetor dos efeitos aleatórios da interação ambientes $\times$ experimentos de dimensão 30×1 ;
- $\tilde{\mathbf{b}}$ é o vetor dos efeitos aleatórios de blocos dentro de experimentos dentro de ambientes, de dimensão 60×1 ;
- $\tilde{\mathbf{p}}$ é o vetor dos efeitos aleatórios de progênies dentro de experimentos, de dimensão 250×1 ;
- $\mathbf{\tilde{p}t}$ é o vetor dos efeitos aleatórios da interação progênies $\times$ linhagens dentro de experimentos, de dimensão 500×1 ;

- $\tilde{\mathbf{t}}\mathbf{a}$ é o vetor dos efeitos aleatórios da interação linhagens $\times$ ambientes dentro de experimentos, de dimensão 60×1 ;
- $\tilde{\mathbf{p}}\mathbf{a}$ é o vetor dos efeitos aleatórios da interação progênes $\times$ ambientes dentro de experimentos, de dimensão 1500×1 ;
- $\tilde{\mathbf{p}}\mathbf{t}\mathbf{a}$ é o vetor dos efeitos aleatórios da interação progênes $\times$ linhagens $\times$ ambientes dentro de experimentos, de dimensão 3000×1 ;

Logo, para obter um vetor $\mathbf{s}$ de dados simulados basta inserir os valores obtidos para os efeitos fixos no vetor $\tilde{\boldsymbol{\beta}}$ e, por meio do comando `rnorm` do software R montar um vetor para cada um dos efeitos aleatórios e depois concatená-los em um único vetor de efeitos aleatórios $\tilde{\boldsymbol{\theta}}$.

O comando para montar o vetor para o efeito de progênes, por exemplo, pode ser dado por: `rnorm(250,0, $\hat{\sigma}_p$ )`, em que $\hat{\sigma}_p$ é o desvio padrão da distribuição do efeito de progênes. Já o comando para montar o vetor da interação entre progênes e linhagens não é tão direto, pois deve ser levada em consideração a restrição $\sum_j pt_{ij(l)} = 0$; uma maneira de montar o vetor da interação progênes por linhagem pode ser dada por:

$$\text{ptl1} = \text{rnorm}(50, 0, \hat{\sigma}_{pt}) \quad (68)$$

$$\text{ptl2} = -\text{ptl1} \quad (69)$$

$$\text{pt} = \text{rbind}(\text{pt11}, \text{pt12}, \text{pt21}, \text{pt22}, \text{pt31}, \text{pt32}, \text{pt41}, \text{pt42}, \text{pt51}, \text{pt52}) \quad (70)$$

Em que `ptl1` representa o vetor dos efeitos aleatórios da interação entre progênes e linhagem 1 dentro do experimento l e `ptl2` representa o vetor dos efeitos aleatórios da interação entre progênes e linhagem 2 dentro do experimento l . Ao forçar que um vetor receba os valores opostos do outro vetor fica confirmada a restrição de que o efeito de linhagens deve se anular dentro de experimentos.

Para obter o vetor dos erros aleatórios $\tilde{\boldsymbol{\varepsilon}}$ basta montar um vetor de dimensões 1000×1 com o comando `rnorm` e com as informações sobre $\hat{\sigma}^2$.

Pode-se notar que todo o lado direito da eq. (59), ou é conhecido, ou pode ser simulado, basta então resolver a equação para obter o vetor $\mathbf{s}$ dos dados simulados.

Serão simulados, no total, mil conjuntos de dados. Para tal basta montar uma matriz $\tilde{\mathbf{B}}$ de dimensões 11×1000 , cujas 1000 colunas sejam todas iguais à $\tilde{\boldsymbol{\beta}}$. Da mesma forma, pode-se montar uma matriz $\tilde{\mathbf{T}}$ de dimensões 5411×1000 cujas colunas sejam simulações independentes do vetor $\tilde{\boldsymbol{\theta}}$, assim como uma matriz $\tilde{\mathbf{E}}$ de dimensões 6000×1000 cujas colunas sejam simulações independentes do vetor $\tilde{\boldsymbol{\varepsilon}}$. Substituindo essas novas matrizes no modelo (59), tem-se:

$$\mathbf{S} = \mathbf{X}\tilde{\mathbf{B}} + \mathbf{Z}\tilde{\mathbf{T}} + \tilde{\mathbf{E}} \quad (71)$$

Sendo assim, a matriz $\mathbf{S}$ de dimensões 6000×1000 possui 1000 colunas, independentes umas das outras, em que, cada coluna representa um conjunto de dados simulados de acordo com as informações obtidas dos dados reais analisados.

Esses 1000 conjuntos de dados simulados foram utilizados na estimação dos componentes de variância estatísticos e genéticos, assim como do grau médio de dominância e coeficiente de herdabilidade. As estimativas, pelo método da ANOVA, foram obtidas com o auxílio do software R, no qual foram montadas as somas de quadrados, quadrados médios e esperanças dos quadrados médios da ANOVA. Já as estimativas relativas ao método REML foram obtidas com auxílio do software SAS, por meio do procedimento PROC MIXED. Os valores das somas de quadrado obtidos pelos software R e SAS foram comparados, a fim de certificar-se que ambos os softwares consideraram o mesmo modelo matemático para os dados.

3.3.1 Comparação dos Métodos de Estimação

Serão obtidos conjuntos de 1000 estimativas para cada componente de variância, variância genética, coeficiente de herdabilidade e grau médio de dominância. Esses conjuntos de estimativas foram utilizados na obtenção de gráficos e estatísticas que auxiliassem na comparação dos dois métodos utilizados. Outros valores muito importantes na comparação dos métodos de estimação são as estimativas obtidas dos dados reais, pois essas são tidas como o valor de referência, aquele que os estimadores deveriam estimar; já que foram esses os valores que foram utilizados na simulação dos dados.

3.3.1.1 Gráficos

Será feito um box plot das 1000 estimativas de cada um dos componentes de variância e variâncias e parâmetros genéticos. Esse gráfico possui informações sobre a média, mediana, quartís e variação dos dados, fornecendo uma ferramenta para a comparação dos métodos de estimação. Por meio da comparação entre box plots das estimativas de um mesmo componente de variância obtidas por ambos os métodos é possível visualizar qual o método que mais se aproximou do valor real em média; qual o método que apresentou um menor erro médio e assim por diante.

3.3.1.2 Estatísticas Descritivas

Serão calculadas as médias, desvios padrão e acurácias relativas (mostra em percentagem o quanto do valor real a estimativa se afasta, quanto menor seu valor melhor é a estimativa) para cada um dos conjuntos de 1000 estimativas, esses valores, juntamente com os box plots, fornecem uma base de comparação para os métodos de estimação, permitindo a escolha do método mais eficiente na estimação dos componentes de variância, variâncias e parâmetros genéticos de interesse.

As equações que fornecem a média, desvio padrão e acurácia relativa são, respectivamente:

$$m = \frac{\sum_z \hat{\alpha}_z}{Z}; \quad (72)$$

$$dp = \sqrt{\frac{\sum_z (\hat{\alpha}_z - m)^2}{\frac{z}{Z - 1}}}; \quad (73)$$

$$AR = \sum_z \left(\frac{\hat{\alpha}_z - \alpha}{\alpha} \right) \frac{100}{Z}. \quad (74)$$

Em que:

- m representa a média de um conjunto de estimativas;

- $\hat{\alpha}_z$ representa a z -ésima estimativa de um determinado parâmetro α , com $z = 1, \dots, 1000$;
- α representa o valor real de determinado parâmetro, aquele utilizado na simulação dos dados;
- Z representa o valor total de estimativas, $Z = 1000$;
- dp representa o desvio padrão de um conjunto de estimativas;
- AR representa a acurácia relativa de um conjunto de estimativas.

4 RESULTADOS E DISCUSSÃO

4.1 Dados Reais

Os dados obtidos para os caracteres produção de grãos (kg/hectare), número de folhas por planta (média por parcela) e altura da planta (média por parcela) foram analisados pelo método da análise da variância (ANOVA) como proposto para o Delineamento III (COMSTOCK e ROBINSON, 1952).

Para o caráter produção de grãos foram obtidos os seguintes resultados:

Tabela 4 - ANOVA para a análise conjunta dos dados de produção de grãos segundo o delineamento fatorial implantado em blocos casualizados, modelo (6), em que CV representa a causa de variação, GL os graus de liberdade, SQ as somas de quadrados e QM os quadrados médios. Sendo que d. significa dentro de

CV	GL	SQ	QM
Ambientes (A)	5	2758428	551685,6
Experimentos (E)	4	12723,41	3180,8525
A×E	20	51566,9	2578,345
Blocos d. A d. E	30	132415,5	4413,85
Progênies (P) d. E	245	330766,4	1350,0669
Linhagens (T) d. E	5	424297,8	84859,56
T×P d. E	245	556316	2270,6775
P×A d. E	1225	488751,4	398,9807
T×A d. E	25	55898,89	2235,9556
P×T×A d. E	1225	460281,2	375,7397
Resíduos	2970	939449,7	316,313
Total	5999	6210895	

Com os valores dos quadrados médios obtidos (Tabela 4) foram calculadas as estimativas dos componentes de variância (método da ANOVA) e dos efeitos fixos do modelo matemático (6).

Tabela 5 - Valores das estimativas dos componentes de variância e efeitos fixos relativos ao modelo (6) para o caráter produção de grãos

Componentes de Variância	Estimativas	Efeitos Fixos	Estimativas
$\hat{\sigma}_a^2$	549,1073	$\hat{\mu}$	116,3521
$\hat{\sigma}_e^2$	-0,2905	$\hat{t}_{11}^2$	8,4207
$\hat{\sigma}_{ae}^2$	-9,5909	$\hat{t}_{12}^2$	-8,4207
$\hat{\sigma}_b^2$	40,9754	$\hat{t}_{21}^2$	6,0891
$\hat{\sigma}_p^2$	39,6286	$\hat{t}_{22}^2$	-6,0891
$\hat{\sigma}_{pt}^2$	157,9115	$\hat{t}_{31}^2$	9,4152
$\hat{\sigma}_{pa}^2$	20,6669	$\hat{t}_{32}^2$	-9,4152
$\hat{\sigma}_{ta}^2$	18,6022	$\hat{t}_{41}^2$	8,3919
$\hat{\sigma}_{pta}^2$	29,7134	$\hat{t}_{42}^2$	-8,3919
$\hat{\sigma}^2$	316,3130	$\hat{t}_{51}^2$	9,2683
		$\hat{t}_{52}^2$	-9,2683

Utilizando os valores dados na Tabela 5 foram obtidas as estimativas para as variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância.

Tabela 6 - Valores das estimativas das variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância para o caráter produção de grãos

Parâmetros Genéticos	Estimativas
$\hat{V}_f$	54,5305
$\hat{V}_a$	158,5144
$\hat{V}_d$	157,9115
$\hat{V}_g$	316,4259
$\hat{V}_{ae}$	82,6677
$\hat{V}_{de}$	29,7134
$\hat{V}_{ge}$	112,3811
$\hat{h}^2$	0,7267
$\hat{GMD}$	1,4115

Os resultados obtidos dessas análises preliminares foram utilizados na simulação de 1000 conjuntos de dados com características semelhantes aos dados originais de produção de grãos.

Os valores dados pela Tabela 5 foram inseridos no programa de simulação, sendo que, no caso das estimativas dos componentes de variância relativos a ambiente ($\hat{\sigma}_a^2$) e a experimento ($\hat{\sigma}_e^2$) foi utilizado o valor nulo nas simulações, já que os valores estimados para esses parâmetros foram negativos. Essa alteração nos componentes de variância relativos ao modelo (6), no entanto, não altera os valores de referência para as variância genéticas (Tabela 6).

Para o caráter altura da planta foram obtidas os seguintes resultados:

Tabela 7 - ANOVA para a análise conjunta dos dados de altura da planta segundo o delineamento fatorial implantado em blocos casualizados, modelo (6), em que CV representa a causa de variação, GL os graus de liberdade, SQ as somas de quadrados e QM os quadrados médios. Sendo que d. significa dentro de

CV	GL	SQ	QM
Ambientes (A)	5	653766,8268	130753,3654
Experimentos (E)	4	5586,339	1396,5847
A×E	20	80664,799	4033,2399
Blocos d. A d. E	30	124364,385	4145,4795
Progenies (P) d. E	245	152196,7892	621,2114
Linhagens (T) d. E	5	43871,7025	8774,3405
T×P d. E	245	101021,5058	412,3327
P×A d. E	1225	157435,4158	128,5187
T×A d. E	25	15402,0625	616,0825
P×T×A d. E	1225	125712,4792	102,6224
Resíduos	2970	291001,115	97,9802
Total	5999	1751023,42	

Utilizando as informações da Tabela 7 foi possível obter estimativas dos componentes de variância e dos efeitos fixos do modelo matemático (6), cujos valores estão apre-

sentados na Tabela 8

Tabela 8 - Valores das estimativas dos componentes de variância e efeitos fixos relativos ao modelo (6) para o caráter altura da planta

Componentes de Variância	Estimativas	Efeitos Fixos	Estimativas
$\hat{\sigma}_a^2$	126,7201	$\hat{\mu}$	202,7081
$\hat{\sigma}_e^2$	-2,6078	$\hat{t}_{11}^2$	2,7695
$\hat{\sigma}_{ae}^2$	-0,7139	$\hat{t}_{12}^2$	-2,7695
$\hat{\sigma}_b^2$	40,4749	$\hat{t}_{21}^2$	1,7696
$\hat{\sigma}_p^2$	20,5389	$\hat{t}_{22}^2$	-1,7696
$\hat{\sigma}_{pt}^2$	25,8092	$\hat{t}_{31}^2$	2,6063
$\hat{\sigma}_{pa}^2$	7,6346	$\hat{t}_{32}^2$	-2,6063
$\hat{\sigma}_{ta}^2$	5,1346	$\hat{t}_{41}^2$	2,9521
$\hat{\sigma}_{pta}^2$	2,3211	$\hat{t}_{42}^2$	-2,9521
$\hat{\sigma}^2$	97,9802	$\hat{t}_{51}^2$	3,1934
		$\hat{t}_{52}^2$	-3,1934

Utilizando os valores fornecidos pela Tabela 8 foi possível calcular estimativas para as variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância, que são dadas na tabela 9:

Assim como para o caráter produção de grãos, foi necessária a alteração dos valores obtidos para $\hat{\sigma}_a^2$ e $\hat{\sigma}_e^2$ (Tabela 7) na programação da simulação de dados semelhantes aos coletados para o caráter altura da planta. Da mesma forma, essa alteração não muda os valores de referência para as variância genéticas dados na Tabela 9.

E finalmente, para o caráter número de folhas tem-se a análise da ANOVA resumida na Tabela 10.

Tabela 9 - Valores das estimativas das variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância para o caráter altura da planta

Parâmetros Genéticos	Estimativas
$\hat{V}_f$	25,2476
$\hat{V}_a$	82,1154
$\hat{V}_d$	25,8092
$\hat{V}_g$	107,9246
$\hat{V}_{ae}$	30,5385
$\hat{V}_{de}$	2,3211
$\hat{V}_{ge}$	32,8597
$\hat{h}^2$	0,8131
$G\hat{M}D$	0,7928

Tabela 10 - ANOVA para a análise conjunta dos dados de número de folhas segundo o delineamento fatorial implantado em blocos casualizados, modelo (6), em que CV representa a causa de variação, GL os graus de liberdade, SQ as somas de quadrados e QM os quadrados médios. Sendo que d. significa dentro de

CV	GL	SQ	QM
Ambientes (A)	5	145,4266	29,0853
Experimentos (E)	4	1,3828	0,3457
A×E	20	6,5934	0,3297
Blocos d. A d. E	30	12,0706	0,4023
Progênies (P) d. E	245	185,305	0,7563
Linhagens (T) d. E	5	6,4095	1,2819
T×P d. E	245	47,3342	0,1932
P×A d. E	1225	141,0556	0,1151
T×A d. E	25	4,0427	0,1617
P×T×A d. E	1225	124,016	0,1012
Resíduos	2970	305,9943	0,103
Total	5999	979,6307	

Utilizando as informações da Tabela 10 foram obtidas as estimativas dos componentes de variância e dos efeitos fixos do modelo matemático (6), dadas na Tabela 11.

Tabela 11 - Valores das estimativas dos componentes de variância e efeitos fixos relativos ao modelo (6) para o caráter número de folhas

Componentes de Variância	Estimativas	Efeitos Fixos	Estimativas
$\hat{\sigma}_a^2$	0,0287	$\hat{\mu}$	5,9663
$\hat{\sigma}_e^2$	-0,0005	$\hat{t}_{11}^2$	2,9679
$\hat{\sigma}_{ae}^2$	-0,0004	$\hat{t}_{12}^2$	-2,9679
$\hat{\sigma}_b^2$	0,003	$\hat{t}_{21}^2$	0,0074
$\hat{\sigma}_p^2$	0,0267	$\hat{t}_{22}^2$	-0,0074
$\hat{\sigma}_{pt}^2$	00,77	$\hat{t}_{31}^2$	0,0342
$\hat{\sigma}_{pa}^2$	0,003	$\hat{t}_{32}^2$	-0,0342
$\hat{\sigma}_{ta}^2$	0,0006	$\hat{t}_{41}^2$	0,0282
$\hat{\sigma}_{pta}^2$	-0,0008	$\hat{t}_{42}^2$	-0,0282
$\hat{\sigma}^2$	0,103	$\hat{t}_{51}^2$	0,0313
		$\hat{t}_{52}^2$	-0,0313

Utilizando os valores fornecidos pela Tabela 11 foram obtidas as estimativas para as variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância, dadas na Tabela 12.

Tabela 12 - Valores das estimativas das variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância para o caráter número de folhas

Parâmetros Genéticos	Estimativas
$\hat{V}_f$	0,0313
$\hat{V}_a$	0,1069
$\hat{V}_d$	0,0077
$\hat{V}_g$	0,1145
$\hat{V}_{ae}$	0,0121
$\hat{V}_{de}$	-0,0009
$\hat{V}_{ge}$	0,0112
$\hat{h}^2$	0,8546
$G\hat{M}D$	0,3787

Já no caso do caráter número de folhas foi necessário utilizar o valor nulo, na simulação de dados, para a estimativa do componente de variância relativo à interação entre progênie, linhagem e ambiente ($\hat{\sigma}_{pta}^2$), além de $\hat{\sigma}_a^2$ e $\hat{\sigma}_e^2$ (Tabela 10), o que causou uma alteração nos valores de referência dos parâmetros genéticos. Para a interação entre as variâncias de dominância e de ambiente ($\hat{V}_{de}$) deve ser considerado o valor nulo e para a interação entre as variâncias genotípica e de ambiente ($\hat{V}_{ge}$) deve ser considerado o mesmo valor dado para a interação entre as variância aditiva e de ambiente ($\hat{V}_{ae}$) que é dado na Tabela 12.

4.2 Dados Simulados

4.2.1 Componetes de Variância

Cada um dos 1000 conjuntos de dados simulados foi analisado pelos métodos da ANOVA e da máxima verossimilhança restrita (REML). Sendo assim, foram obtidas 1000 estimativas, por método, para cada um dos componentes de variância do modelo (6).

Para cada conjunto de estimativas foram calculadas estatísticas descritivas (média, desvio-padrão e acurácia relativa); foi feito, também, um box-plot. Os resultados obtidos foram separados por componente de variância, ou seja, num mesmo gráfico, ou tabela, são expostos os resultados de um mesmo componente de variância considerando os três dife-

rentes caracteres de milho estudados.

Os resultados obtidos para o componente de variância relativo ao efeito de ambiente (σ_a^2) são dados na Tabela 13

Tabela 13 - Estatísticas descritivas das 1000 estimativas de σ_a^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_a^2$	Produção	ANOVA	549,1073	562,935	352,8557	50,1207%
$\hat{\sigma}_a^2$	Produção	REML	549,1073	558,7159	368,5159	52,6473%
$\hat{\sigma}_a^2$	Altura	ANOVA	126,7201	125,8927	81,9451	50,2936%
$\hat{\sigma}_a^2$	Altura	REML	126,7201	125,1741	82,2207	50,687%
$\hat{\sigma}_a^2$	Folhas	ANOVA	0,0287	0,0287	0,0187	51,1281%
$\hat{\sigma}_a^2$	Folhas	REML	0,0287	0,0285	0,0187	51,2633%

Pode-se notar, tanto pela Tabela 13 quanto pela Figura 7, que ambos os métodos estimaram σ_a^2 de maneira praticamente igual, mesmo sendo os valores de referência bastante diferentes para cada um dos caracteres simulados. Em média ambos os métodos forneceram valores muito próximos do valor real, sendo que conforme diminui o valor de referência mais os métodos se aproximaram. Pode-se notar, também, que os desvios-padrão mantiveram uma proporcionalidade de aproximadamente $\frac{2}{3}$ em relação à média para todos os diferentes caracteres, o que causou uma semelhança nos valores obtidos para a acurácia relativa. Considerando todas as informações pode-se dizer que ambos os métodos são similares na estimação de σ_a^2 , porém ambos podem não ser muito eficientes, já que apresentaram uma média de erro na grandeza de cerca de 50% do valor que se pretende estimar, sendo interessante obter diversas estimativas para um mesmo componente, e utilizar a média dessas estimativas como valor final.

Figura 7 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_a^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Os resultados para o componente de variância relativo ao efeito de experimentos (σ_e^2) são dados na Tabela 14.

Tabela 14 - Estatísticas descritivas das 1000 estimativas de σ_e^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_e^2$	Produção	ANOVA	-	0,0274	5,1064	...
$\hat{\sigma}_e^2$	Produção	REML	-	-	-	...
$\hat{\sigma}_e^2$	Altura	ANOVA	-	0,4374	3,8366	...
$\hat{\sigma}_e^2$	Altura	REML	-	-	-	...
$\hat{\sigma}_e^2$	Folhas	ANOVA	-	-	0,0014	...
$\hat{\sigma}_e^2$	Folhas	REML	-	-	-	...

Nota: Sinais convencionais utilizados:

- Dado numérico igual a zero não resultante de arredondamento.
- ... Dado numérico não disponível.

No caso de σ_e^2 fica claro, ao observar a Tabela 14 e a Figura 8 que os métodos da ANOVA e da REML não se igualam na estimação. Esse é um caso especial, pois o valor de referência é nulo, logo, era esperado que o método da ANOVA fornecesse, tanto estimativas positivas quanto negativas, se aproximando, média, do valor nulo; e que o método da REML fosse truncado em zero, por não aceitar estimativas negativas. O método da ANOVA atuou de acordo com sua descrição, foi não viesado, apresentando cerca de 50% das estimativas negativas e em média aproximou-se bastante do valor de referência; porém o método da ANOVA apresentou um desvio-padrão relativamente alto nos casos dos caracteres de produção de grãos e da altura da planta. O método da REML, no entanto, foi perfeito, estimando corretamente o valor de referência para cada um dos conjuntos simulados, considerando todos os caracteres.

Figura 8 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_e^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Considerando o componente de variância relativo ao efeito da interação entre ambientes e experimentos (σ_{ae}^2) têm-se os resultados dados na Tabela 15.

Tabela 15 - Estatísticas descritivas das 1000 estimativas de σ_{ae}^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_{ae}^2$	Produção	ANOVA	-	9,8809	12,2114	...
$\hat{\sigma}_{ae}^2$	Produção	REML	-	4,7238	7,3989	...
$\hat{\sigma}_{ae}^2$	Altura	ANOVA	-	3,7183	9,8648	...
$\hat{\sigma}_{ae}^2$	Altura	REML	-	3,7941	6,1068	...
$\hat{\sigma}_{ae}^2$	Folhas	ANOVA	-	0,0003	0,0027	...
$\hat{\sigma}_{ae}^2$	Folhas	REML	-	0,0011	0,0019	...

Nota: Sinais convencionais utilizados:

- Dado numérico igual a zero não resultante de arredondamento.
- ... Dado numérico não disponível.

Assim como para σ_e^2 , o valor de referência para σ_{ae}^2 é zero, novamente o método da REML foi o mais eficiente na estimação do componente, apresentando médias mais próximas do valor de referência (exceto pro caso do caráter altura da planta, no qual as médias são praticamente as mesmas) e desvios-padrão menores do que os obtidos pelo método da ANOVA (Tabela 15). Novamente, o método da ANOVA obteve estimativas negativas, porém mostrou viés na estimação (Figura 9), quanto maior a grandeza dos dados observados maior o viés. O viés apresentado pelo método da REML, além de esperado, se manteve relativamente constante para os diferentes caracteres, como pode ser visto ao observar a proporção entre os desvios-padrão e médias obtidos (variando entre $\frac{2}{3}$ a $\frac{4}{7}$).

Figura 9 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_{ae}^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

A seguir, na Tabela 16, são dados os resultados obtidos para o componente de variância relativo ao efeito de bloco (σ_b^2):

Tabela 16 - Estatísticas descritivas das 1000 estimativas de σ_b^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_b^2$	Produção	ANOVA	40,9754	41,3714	11,7188	22,5614%
$\hat{\sigma}_b^2$	Produção	REML	40,9754	39,5655	10,2974	20,4508%
$\hat{\sigma}_b^2$	Altura	ANOVA	40,4749	40,8819	10,6905	20,5805%
$\hat{\sigma}_b^2$	Altura	REML	40,4749	38,412	9,0121	18,1479%
$\hat{\sigma}_b^2$	Folhas	ANOVA	0,0029	0,003	0,001	27,4526%
$\hat{\sigma}_b^2$	Folhas	REML	0,0029	0,0029	0,001	26,5501%

Assim como para σ_a^2 , ambos os métodos de estimação foram similares na predição de σ_b^2 (Tabela 16 e Figura 10). As estimativas variaram pouco para os caracteres produção de grãos e altura da planta, o que é compatível com o fato dos valores de referência serem quase os mesmos. Apesar dos valores obtidos para a acurácia relativa não serem tão baixos quanto desejável eles fornecem a informação de que os métodos erraram em média entre $\frac{1}{5}$ e $\frac{1}{4}$ do valor de referência que deveriam estimar, ou seja, apesar de não serem perfeitos, ambos os métodos foram satisfatórios na estimativa de σ_b^2 .

Figura 10 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_b^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Os resultados obtidos para o componente de variância relativo ao efeito de progênie (σ_p^2), um dos componentes mais importantes para o estudo das características genéticas, são apresentados na Tabela 17.

Tabela 17 - Estatísticas descritivas das 1000 estimativas de σ_p^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_p^2$	Produção	ANOVA	39,6286	39,6662	5,4502	10,8914%
$\hat{\sigma}_p^2$	Produção	REML	39,6286	-	-	100%
$\hat{\sigma}_p^2$	Altura	ANOVA	20,5289	20,5154	2,4157	9,3886%
$\hat{\sigma}_p^2$	Altura	REML	20,5289	0,1101	0,5344	99,4638%
$\hat{\sigma}_p^2$	Folhas	ANOVA	0,0267	0,0268	0,0079	23,7165%
$\hat{\sigma}_p^2$	Folhas	REML	0,0267	0,0194	0,0079	33,2439%

Nota: Sinal convencional utilizado:

- Dado numérico igual a zero não resultante de arredondamento.

Fica claro ao observar a Figura 11 que existe uma grande diferença entre os métodos de estimação da ANOVA e da REML na predição de σ_p^2 . Sendo que o método da ANOVA é bem mais eficiente na estimação desse componente. Levando em consideração as estatísticas fornecidas na Tabela 17 pode-se notar que quanto menor o valor de σ_p^2 menor o erro resultante da estimação pelo método da REML. Um dos motivos possíveis para tamanha disparidade entre os métodos de estimação é o fato do método da ANOVA levar em consideração as restrições (eq. (7), (8) e (9)) além da usualmente utilizada para efeitos fixos ($\sum_j t_{j(l)} = 0$), dadas no item (4.2.1.1), impostas ao modelo matemático (6) na simulação dos dados. No caso do método REML a inserção das mesmas restrições no cálculo das estimativas é mais complexa, sendo necessária a formulação de uma matriz de variâncias e covariâncias que respeitasse todas as restrições do modelo (6). Existem no pacote PROC MIXED do software SAS diversas matrizes de variâncias e covariâncias, porém nenhuma delas leva em consideração todas as restrições aqui aplicadas, sendo que a matriz que melhor se ajusta é a de componentes de variância, que leva em consideração apenas a restrição $\sum_j t_{j(l)} = 0$.

Figura 11 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_p^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Outro componente bastante importante no estudo dos parâmetros genéticos dos caracteres de interesse é o componente de variância relativo ao efeito da interação entre progênie e linhagem (σ_{pt}^2), os resultados obtidos para esse componente podem ser vistos na Tabela 18.

Tabela 18 - Estatísticas descritivas das 1000 estimativas de σ_{pt}^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_{pt}^2$	Produção	ANOVA	157,9115	317,6421	31,7773	101,152%
$\hat{\sigma}_{pt}^2$	Produção	REML	157,9115	197,5939	16,0948	25,1445%
$\hat{\sigma}_{pt}^2$	Altura	ANOVA	25,8092	51,8207	5,4201	100,7841%
$\hat{\sigma}_{pt}^2$	Altura	REML	25,8092	46,4735	3,6702	80,0657%
$\hat{\sigma}_{pt}^2$	Folhas	ANOVA	0,0077	0,0152	0,0021	98,7932%
$\hat{\sigma}_{pt}^2$	Folhas	REML	0,0077	0,0151	0,0021	97,1923%

Ao observar a Tabela 18 e a Figura 12 pode-se perceber que ambos os métodos se diferenciaram na predição de σ_{pt}^2 para os três caracteres estudados. O que se manteve constante foi a superioridade do método da REML, o qual apresentou, em média, estimativas mais próximas do valor de referência, além de fornecer desvios-padrão e acurácias relativas menores, e o fato de que ambos os métodos de estimação fornecem estimativas superestimadas quase que exclusivamente. Pode-se notar que quanto menor o valor de referência mais os métodos se aproximam entre si, porém essa aproximação entre os métodos resultou em uma estimação ineficaz para ambos. Aparentemente ambos os métodos são ineficazes na estimação de σ_{pt}^2 , sendo que o método que menos erra é o da REML.

Figura 12 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_{pt}^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Considerando o componente de variância relativo à interação entre progênie e ambiente (σ_{pa}^2) têm-se os resultados fornecidos na Tabela 19.

Tabela 19 - Estatísticas descritivas das 1000 estimativas de σ_{pa}^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_{pa}^2$	Produção	ANOVA	20,6669	35,6684	5,1674	72,5952%
$\hat{\sigma}_{pa}^2$	Produção	REML	20,6669	0,045	0,4722	99,7821%
$\hat{\sigma}_{pa}^2$	Altura	ANOVA	7,6346	7,6747	1,4334	15,0786%
$\hat{\sigma}_{pa}^2$	Altura	REML	7,6346	5,1073	1,6632	33,9486%
$\hat{\sigma}_{pa}^2$	Folhas	ANOVA	0,003	0,3033	0,0129	9912,524%
$\hat{\sigma}_{pa}^2$	Folhas	REML	0,003	0,3025	0,0129	9885,419%

Assim como para o componente σ_{pt}^2 ambos os métodos de estimação forneceram resultados diferentes para cada um dos caracteres (Figura 13), sendo que o método da ANOVA foi o que forneceu estimativas satisfatórias, porém, apenas para o caso do caráter altura da planta (Tabela 19). Além dos métodos se diferenciarem para cada caráter, eles também forneceram estimativas bastante distintas entre si: para o caráter produção de grãos o método da ANOVA forneceu, quase que exclusivamente, estimativas superestimadas; já o método da REML forneceu apenas estimativas subestimadas. Para o caráter número de folhas ambos os métodos de estimação superestimaram σ_{pa}^2 em todas as estimativas. Aparentemente ambos os métodos são razoavelmente confiáveis apenas para um intervalo pequeno de valores, porém seriam necessárias outras simulações com outros valores de referência para verificar tal suposição.

Figura 13 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_{pa}^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

A seguir, na Tabela 20, são dados os resultados referentes ao componente de variância relativo ao efeito da interação entre linhagem e ambiente (σ_{ta}^2):

Tabela 20 - Estatísticas descritivas das 1000 estimativas de σ_{ta}^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_{ta}^2$	Produção	ANOVA	18,6022	18,6826	7,7842	32,6859%
$\hat{\sigma}_{ta}^2$	Produção	REML	18,6022	17,71842	7,1003	31,0462%
$\hat{\sigma}_{ta}^2$	Altura	ANOVA	5,1346	7,9215	3,3512	64,5401%
$\hat{\sigma}_{ta}^2$	Altura	REML	5,1346	7,7774	3,2408	61,913%
$\hat{\sigma}_{ta}^2$	Folhas	ANOVA	0,0006	0,0006	0,0005	65,5335%
$\hat{\sigma}_{ta}^2$	Folhas	REML	0,0006	0,0006	0,0005	61,5756%

Para σ_{ta}^2 ambos os métodos de estimação mostram certa consistência dependendo do valor de referência do componente (Tabela 20 e Figura 14). Em todos os casos ambos os métodos foram bastante semelhantes entre si, sendo que o método da REML apresentou uma acurácia relativa média um pouco menor do que a obtida para as estimativas pelo método da ANOVA. Aparentemente, a eficiência dos métodos está relacionada ao valor de σ_{ta}^2 ; para o caráter produção de grãos (maior valor de referência) ambos os métodos apresentaram uma taxa de erro duas vezes menor do que as taxas obtidas para os outros dois caracteres.

Figura 14 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_{ta}^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Para o componente de variância relativo ao efeito da interação entre progênie, linhagem e ambiente (σ_{pta}^2) foram obtidos os seguintes resultados, dados na Tabela 21:

Tabela 21 - Estatísticas descritivas das 1000 estimativas de σ_{pta}^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}_{pta}^2$	Produção	ANOVA	29,7134	29,224	8,7601	23,4048%
$\hat{\sigma}_{pta}^2$	Produção	REML	29,7134	50,2254	7,4969	69,0938%
$\hat{\sigma}_{pta}^2$	Altura	ANOVA	2,3211	4,7009	2,5555	122,0328%
$\hat{\sigma}_{pta}^2$	Altura	REML	2,3211	4,873	2,5169	125,4644%
$\hat{\sigma}_{pta}^2$	Folhas	ANOVA	-	-	0,0025	...
$\hat{\sigma}_{pta}^2$	Folhas	REML	-	0,0009	0,0014	...

Nota: Sinais convencionais utilizados:

- Dado numérico igual a zero não resultante de arredondamento.
- ... Dado numérico não disponível.

Pode-se notar, pela Tabela 21 e Figura 15, que o método da REML forneceu, em média, estimativas superestimadas para σ_{pta}^2 para todos os caracteres estudados. Já o método da ANOVA forneceu, em média, estimativas superestimadas apenas no caso do caráter altura da planta, sendo que, para esse caráter, ambos os métodos apresentaram uma acurácia relativa muito alta, estimando em média, valores superiores a duas vezes o valor de referência. No caso dos outros dois caracteres o método da ANOVA se mostrou mais ou tão eficiente quanto o método da REML.

Figura 15 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ_{pta}^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Finalmente, os resultados obtidos para o componente de variância relativo ao efeito do erro aleatório σ^2 são dados na Tabela 22.

Tabela 22 - Estatísticas descritivas das 1000 estimativas de σ^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: CV representa o componente de variância estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

CV	Caráter	Método	VR	Média	DP	AR
$\hat{\sigma}^2$	Produção	ANOVA	316,3130	316,3028	8,4744	2,1406%
$\hat{\sigma}^2$	Produção	REML	316,3130	316,3992	8,3802	2,1116%
$\hat{\sigma}^2$	Altura	ANOVA	97,9802	98,0064	2,5445	2,0718%
$\hat{\sigma}^2$	Altura	REML	97,9802	97,9984	2,5298	2,064%
$\hat{\sigma}^2$	Folhas	ANOVA	0,103	0,1031	0,0027	2,1029%
$\hat{\sigma}^2$	Folhas	REML	0,103	0,1025	0,0024	1,9159%

Ambos os métodos de estimação foram extremamente eficientes na estimação de σ^2 , como pode ser visto pelas Tabela 22 e Figura 16, em média as estimativas obtidas por ambos os métodos foram muito próximas do valor real e os desvios-padrão foram, relativamente, os mais baixos fornecidos por ambos os métodos. Esses dois fatores combinados implicaram em valores muito baixos para a acurácia relativa, ou seja, a taxa de erro na predição de σ^2 é mínima, independentemente do valor de referência e do método utilizado.

Figura 16 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para σ^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

4.2.2 Parâmetros genéticos

Os conjuntos de estimativas dos componentes da variância relativos ao modelo 6 foram utilizados no cálculo das variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância para os três caracteres estudados.

Assim, como para os componentes de variância relativos ao modelo matemático, foram calculadas estatísticas descritivas (média, desvio-padrão e acurácia relativa) e foi feito um box-plot. Da mesma maneira, os resultados são apresentados de acordo com o parâmetro ao qual são relativos, contendo as informações obtidas de todos os caracteres em uma única tabela e gráfico.

Os resultados obtidos para a variância aditiva (V_a) são dados na Tabela 23.

Tabela 23 - Estatísticas descritivas das 1000 estimativas de V_a obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{V}_a$	Produção	ANOVA	158,5144	158,6649	21,8008	10,8914%
$\hat{V}_a$	Produção	REML	158,5144	-	-	100%
$\hat{V}_a$	Altura	ANOVA	82,1154	82,0617	9,6628	9,3886%
$\hat{V}_a$	Altura	REML	82,1154	0,4403	2,1375	99,4638%
$\hat{V}_a$	Folhas	ANOVA	0,1068	0,1074	0,0315	23,7165%
$\hat{V}_a$	Folhas	REML	0,1068	0,0775	0,0318	33,2439%

Nota: Sinal convencional utilizado:

- Dado numérico igual a zero não resultante de arredondamento.

Os resultados obtidos para V_a são relacionados com aqueles obtidos para σ_p^2 , já que $\hat{V}_a = 4\hat{\sigma}_p^2$. Ao observar as Figuras 17 e 11, e as Tabelas 23 e 17, pode-se notar que, a não ser pela escala, as relações entre os gráficos e as estatísticas descritivas são as mesmas. Sendo assim, pelos mesmos motivos já discutidos, pode-se dizer que o método da ANOVA foi o mais eficiente na estimação de V_a .

Figura 17 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para V_a para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Os resultados para a variância de dominância (V_d) são dados na Tabela 24.

Tabela 24 - Estatísticas descritivas das 1000 estimativas de V_d obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{V}_d$	Produção	ANOVA	157,9115	317,6421	31,7773	101,152%
$\hat{V}_d$	Produção	REML	157,9115	197,5939	16,0948	25,1445%
$\hat{V}_d$	Altura	ANOVA	25,8092	51,8207	5,4201	100,7841%
$\hat{V}_d$	Altura	REML	25,8092	46,4735	3,6702	80,0657%
$\hat{V}_d$	Folhas	ANOVA	0,0077	0,0152	0,0021	98,7932%
$\hat{V}_d$	Folhas	REML	0,0077	0,0151	0,0021	97,1923%

Assim como para V_a os resultados obtidos para V_d são relacionados aos obtidos para σ_{pt}^2 , já que $\hat{V}_d = \hat{\sigma}_{pt}^2$. Pode-se notar que as Figuras 18 e 12, e as Tabelas 24 e 18, são idênticas. Sendo assim, pelos mesmos motivos já discutidos, tem-se que nenhum dos métodos é eficiente na estimação de V_d , sendo que o que menos errou foi o método da REML.

Figura 18 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para V_d para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Considerando a variância genotípica (V_g) têm-se os resultados dados na Tabela

Tabela 25 - Estatísticas descritivas das 1000 estimativas de V_g obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{V}_g$	Produção	ANOVA	316,4259	476,307	37,5590	50,5272%
$\hat{V}_g$	Produção	REML	316,4259	197,5939	16,0948	25,1445%
$\hat{V}_g$	Altura	ANOVA	107,9246	133,8824	11,0621	24,0696%
$\hat{V}_g$	Altura	REML	107,9246	46,9138	3,9598	56,5309%
$\hat{V}_g$	Folhas	ANOVA	0,1145	0,1226	0,0315	22,7664%
$\hat{V}_g$	Folhas	REML	0,1145	0,0926	0,0315	27,0107%

As estimativas de V_g são obtidas pela soma das estimativas de V_a e V_d , sendo assim os resultados obtidos para V_g são uma combinação daqueles obtidos para as duas variâncias genéticas estudadas no início desse item. Ao observar a Tabela 25 e a Figura 19 pode-se perceber que a eficiência dos métodos de estimativa na predição de V_g depende do caráter ao qual V_g se refere, porém ambos os métodos mantêm determinadas características na predição de V_g . Considerando os caracteres de produção de grãos e altura da planta o método da ANOVA forneceu, exclusivamente, estimativas superestimadas; enquanto o método da REML forneceu, também exclusivamente, estimativas subestimadas. Além disso, ambos os métodos mantiveram a mesma relação de proporção entre as médias e desvios-padrão. Sendo assim, o fator principal que fez um método errar mais, ou menos, foi proximidade entre as médias das estimativas e o valor de referência. Algo semelhante ocorreu para o caráter número de folhas, porém nesse caso, os métodos foram menos viesados, apesar de manterem, em média, a superestimação (ANOVA) ou subestimação (REML) dos dados. Além disso, ambos os métodos apresentaram o mesmo desvio-padrão, e uma acurácia relativa semelhante, porém o método que se sobressaiu para esse caráter foi o da ANOVA. Não é possível, com base nas informações aqui obtidas, apontar um dos métodos como o mais eficiente na predição de V_g .

Figura 19 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para V_g para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

A seguir (Tabela 26) são dados os resultados obtidos para a interação entre as variâncias aditiva e de ambiente (V_{ae}):

Tabela 26 - Estatísticas descritivas das 1000 estimativas de V_{ae} obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{V}_{ae}$	Produção	ANOVA	82,6677	142,6735	20,6697	72,5952%
$\hat{V}_{ae}$	Produção	REML	82,6677	0,1801	1,8887	99,7821%
$\hat{V}_{ae}$	Altura	ANOVA	30,5385	30,6988	5,7338	15,0786%
$\hat{V}_{ae}$	Altura	REML	30,5385	20,4291	6,6529	33,9486%
$\hat{V}_{ae}$	Folhas	ANOVA	0,0121	1,2134	0,0519	9912,524%
$\hat{V}_{ae}$	Folhas	REML	0,0121	1,2101	0,0516	9885,419%

Os resultados obtidos para V_{ae} são relacionados com aqueles obtidos para σ_{pa}^2 , já que $\hat{V}_{ae} = 4\hat{\sigma}_{pa}^2$. Ao observar as Figuras 20 e 13, e as Tabelas 26 e 19, pode-se notar que, a não ser pela escala, as relações entre os gráficos e as estatísticas descritivas são as mesmas. Sendo assim, pelos mesmos motivos já discutidos, não é possível tirar uma conclusão a respeito dessa variância apenas com os casos aqui estudados. Pode-se supor a existência de um curto intervalo de valores para os quais os métodos seriam satisfatórios, com o método da ANOVA sendo o mais indicado, porém se faz necessário o estudo de outras simulações a fim de tirar uma conclusão a esse respeito.

Figura 20 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para V_{ae} para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Os resultados obtidos para a interação entre as variâncias de dominância e de ambiente (V_{de}), são dados na Tabela 27.

Tabela 27 - Estatísticas descritivas das 1000 estimativas de V_{de} obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{V}_{de}$	Produção	ANOVA	29,7134	29,224	8,7601	23,4048%
$\hat{V}_{de}$	Produção	REML	29,7134	50,2254	7,4969	69,0938%
$\hat{V}_{de}$	Altura	ANOVA	2,3211	4,7009	2,5555	122,0328%
$\hat{V}_{de}$	Altura	REML	2,3211	4,873	2,5169	125,4644%
$\hat{V}_{de}$	Folhas	ANOVA	-	0	0,0025	...
$\hat{V}_{de}$	Folhas	REML	-	0,0009	0,0014	...

Nota: Sinais convencionais utilizados:

- Dado numérico igual a zero não resultante de arredondamento.
- ... Dado numérico não disponível.

Assim como para V_{ae} os resultados obtidos para V_{de} são relacionados aos obtidos para σ_{pta}^2 , já que $\hat{V}_{de} = \hat{\sigma}_{pta}^2$. Pode-se notar que as Figuras 21 e 15, e as Tabelas 27 e 21, são idênticas. Sendo assim, pelos mesmos motivos já discutidos, tem-se que os métodos variam quanto a eficiência na predição de V_{de} , sendo que o o método da ANOVA se mostrou mais eficiente, ou no mínimo, tão eficiente quanto o método da REML.

Figura 21 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para V_{de} para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Considerando a interação entre as variâncias genotípica e de ambiente (V_{ge}) têm-se os seguintes resultados (Tabela 28):

Tabela 28 - Estatísticas descritivas das 1000 estimativas de V_{ge} obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{V}_{ge}$	Produção	ANOVA	112,3811	171,8975	23,8642	53,0506%
$\hat{V}_{ge}$	Produção	REML	112,3811	50,4056	7,6133	55,1476%
$\hat{V}_{ge}$	Altura	ANOVA	32,8597	35,3997	6,8192	17,5763%
$\hat{V}_{ge}$	Altura	REML	32,8597	25,3021	5,6249	24,2162%
$\hat{V}_{ge}$	Folhas	ANOVA	0,0121	1,2134	0,052	9912,24%
$\hat{V}_{ge}$	Folhas	REML	0,0121	1,2111	0,0516	9893,575%

As estimativas de V_{ge} são obtidas pela soma das estimativas de V_{ae} e V_{de} . Sendo assim, os resultados obtidos para V_{ge} são uma combinação daqueles obtidos para V_{ae} e V_{de} . Ao observar a Tabela 28 e a Figura 22 pode-se perceber que a eficiência dos métodos de estimativa na predição de V_{ge} depende do caráter ao qual ele se refere, porém ambos os métodos mantêm determinadas características na predição de V_{ge} . A proporção entre médias e desvios-padrão são semelhantes entre métodos, embora sejam distintas entre caracteres. Sendo assim, a eficiência dos métodos é dependente da proximidade média entre as estimativas fornecidas e o valor de referência. Pode-se notar, também, que o método da REML tende a se aproximar do método da ANOVA, conforme diminui o valor de referência estudado. Para o caso do caráter produção de grãos o método da REML forneceu, apenas, estimativas subestimadas; para o caráter altura da planta o método da REML continuou a fornecer estimativas, em média, subestimadas, porém não exclusivamente e com uma acurácia relativa bem menor; já para o caráter número de folhas o método da REML forneceu, exclusivamente, estimativas superestimadas, obtendo estimativas muito semelhantes àquelas obtidas pelo método da ANOVA. O método da ANOVA se mostrou mais eficiente na predição de V_{ge} , embora tenha sido satisfatório apenas no caso do caráter altura da planta.

Figura 22 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para V_{ge} para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

A seguir (Tabela 29) são dados os resultados referentes a variância fenotípica (V_f):

Tabela 29 - Estatísticas descritivas das 1000 estimativas de V_f obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; e AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{V}_f$	Produção	ANOVA	54,5305	55,8179	5,3907	7,9745%
$\hat{V}_f$	Produção	REML	54,5305	13,1871	0,3513	75,8171%
$\hat{V}_f$	Altura	ANOVA	25,2476	25,2386	2,4086	7,6079%
$\hat{V}_f$	Altura	REML	25,2476	4,6189	0,5594	81,7054%
$\hat{V}_f$	Folhas	ANOVA	0,0313	0,0564	0,0076	80,4815%
$\hat{V}_f$	Folhas	REML	0,0313	0,0488	0,0076	56,2895%

A estimação de V_f depende dos valores obtidos para $\hat{\sigma}_p^2$, $\hat{\sigma}_{pa}^2$ e $\hat{\sigma}^2$. Tem-se que ambos os métodos de estimação foram extremamente eficientes na predição de σ^2 , sendo assim, esse componente pode ser considerado como sem efeito na comparação entre os métodos quanto à estimação de V_f . Ao estudar a Tabela 29 e a Figura 23 nota-se que, para os caracteres de produção de grãos e altura da planta, o método da ANOVA foi muito mais eficaz na predição de V_f do que o método da REML, isso se deve ao fato do método da ANOVA ter sido mais eficiente do que o método da REML nas predições de σ_p^2 e σ_{pa}^2 para ambos os caracteres. Para o caráter número de folhas o método da REML foi melhor do que o da ANOVA, porém apresentou uma acurácia relativa de mais de 50%. A maior eficiência do método da REML na predição de V_f para o caráter número de folhas se deve ao fato desse método ter subestimado σ_p^2 e superestimado σ_{pa}^2 , obtendo, então, um valor mais próximo do real; o mesmo não ocorre para o método da ANOVA, pois apesar de também superestimar σ_{pa}^2 ele forneceu estimativas não viesadas para σ_p^2 . Sendo assim, em média, as estimativas de V_f também foram superestimadas.

Figura 23 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para V_f para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Para o coeficiente de herdabilidade h^2 foram obtidos os resultados mostrados na Tabela 30.

Tabela 30 - Estatísticas descritivas das 1000 estimativas de h^2 obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$\hat{h}^2$	Produção	ANOVA	0,7267	0,7078	0,0304	3,7932%
$\hat{h}^2$	Produção	REML	0,7267	-	-	100%
$\hat{h}^2$	Altura	ANOVA	0,8131	0,8111	0,0189	1,8491%
$\hat{h}^2$	Altura	REML	0,8131	0,016	0,0694	98,0295%
$\hat{h}^2$	Folhas	ANOVA	0,8546	0,4653	0,0819	45,5491%
$\hat{h}^2$	Folhas	REML	0,8546	0,3801	0,1097	55,5279%

Nota: Sinal convencional utilizado:

— Dado numérico igual a zero não resultante de arredondamento.

Assim como para V_f a predição de h^2 depende dos valores obtidos para $\hat{\sigma}_p^2$, $\hat{\sigma}_{pa}^2$ e $\hat{\sigma}^2$. Porém para h^2 os valores obtidos para $\hat{\sigma}_p^2$ têm um peso maior do que para V_f . Sendo assim, o método da ANOVA foi mais eficiente na predição de h^2 para todos os caracteres estudados, Tabela 30 e Figura 24, sendo pouco satisfatório apenas para o caráter número de folhas.

Figura 24 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para h^2 para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

Finalmente, os resultados obtidos para o grau médio de dominância GMD são dados na Tabela 31.

Tabela 31 - Estatísticas descritivas das 1000 estimativas de GMD obtidas dos dados simulados pelos métodos da ANOVA e REML, para os caracteres produção de grãos, altura da planta e número de folhas. Em que: PG representa o parâmetro genético estimado; VR o valor de referência do componente; DP o desvio-padrão das estimativas; AR a acurácia média

PG	Caráter	Método	VR	Média	DP	AR
$G\hat{M}D$	Produção	ANOVA	1,4115	2,0133	0,1777	42,6322%
$G\hat{M}D$	Produção	REML	1,4115	...	...	...
$G\hat{M}D$	Altura	ANOVA	0,7928	1,1282	0,0899	42,2946%
$G\hat{M}D$	Altura	REML	0,7928	6,3751	6,0487	143,4937
$G\hat{M}D$	Folhas	ANOVA	0,3787	0,5534	0,1148	46,205%
$G\hat{M}D$	Folhas	REML	0,3787	0,6872	0,2695	81,4872%

Nota: Sinal convencional utilizado:

... Dado numérico não disponível.

A predição do GMD depende das predições de V_a e V_d , em que $G\hat{M}D = \sqrt{\frac{2V_d}{V_a}}$, ou seja, não é possível estimar GMD se a estimativa de V_a for nula, o que ocorreu para todos os caracteres durante a estimação pelo método da REML. Não foi possível obter estimativas de GMD , pelo método da REML, para o caráter produção de grãos, pois todas as estimativas de V_a foram nulas (Tabela 23). Para o Caráter altura da planta foi possível obter apenas 73 estimativas para GMD (utilizando os valores discrepantes mostrados na Figura 17). O maior número de estimativas de GMD fornecidas pelo método da REML foi para o caráter número de folhas (995 estimativas). Por serem conjuntos de tamanhos diferentes os gráficos de box-plot foram feitos separadamente para cada método e cada caráter (Figura 25). Em contraste com o método da REML o método da ANOVA foi bastante consistente na estimação de GMD , como pode ser visto na Tabela 31 e Figura 25. Nota-se que o método da ANOVA superestimou, em média, o GMD em quase metade do seu valor para todos os caracteres, apresentando desvios-padrão bastante baixos; já para o método da REML, quando forneceu estimativas para GMD , essas se mostraram bastante superestimadas e com altos desvios-padrão. Sendo assim, o método mais indicado para a predição do GMD é o método da

ANOVA. Deve-se estar ciente, no entanto, de que as estimativas serão, muito provavelmente, superestimadas.

Figura 25 - Boxplots dos conjuntos de 1000 estimativas encontradas pelos métodos da ANOVA e REML para *GMD* para os caracteres produção de grãos, altura da planta e número de folhas; em que $\square$ representa o valor de referência (aquele que os estimadores deveriam estimar) e $\bullet$ representa a média das estimativas

5 CONCLUSÕES

5.1 Componentes de Variância

Os métodos da ANOVA e da REML se mostraram semelhantes na predição dos componentes de variância relativos aos efeitos de ambiente (σ_a^2), de bloco (σ_b^2), do erro aleatório (σ^2) e da interação entre linhagem e ambiente (σ_{ta}^2). Sendo que ambos os métodos se mostraram muito satisfatórios na predição de σ^2 ; satisfatórios na predição de σ_b^2 ; variaram entre satisfatório e pouco satisfatório na predição de σ_{ta}^2 , dependendo do valor de referência do componente; e foram pouco satisfatórios na predição de σ_a^2 .

O método da ANOVA se mostrou: satisfatório e mais eficiente do que o método da REML na predição do componente de variância relativo ao efeito de progênie (σ_p^2); pouco satisfatório, porém mais eficiente, ou no mínimo tão eficiente quanto o método da REML na predição dos componentes de variância relativos aos efeitos da interação entre progênie e ambiente (σ_{pa}^2) e da interação entre progênie, linhagem e ambiente (σ_{pta}^2).

O método da REML se mostrou: bastante satisfatório e mais eficiente do que o método da ANOVA na predição dos componentes de variância relacionados aos efeitos de experimento (σ_e^2) e da interação entre ambiente e experimento (σ_{ae}^2); e pouco satisfatório, porém mais eficiente, ou no mínimo tão eficiente quanto o método da ANOVA na predição do componente de variância relativo ao efeito da interação entre progênie e linhagem (σ_{pt}^2).

No geral, ambos os métodos foram semelhantes nas predições dos componentes de variância do modelo 6, cada um apresentando vantagens e desvantagens. No entanto, não pode se pensar somente no modelo matemático indicado para estudar o delineamento estatístico utilizado, deve se levar em consideração o fato do experimento ter sido implantado respeitando um delineamento genético, além do estatístico. Sendo assim, acredita-se que o maior interesse do pesquisador está no estudo dos componentes de variância do modelo que serão utilizados no estudo das características genéticas da variável resposta. Nesse caso, o método mais indicado é o da ANOVA, já que foi muito superior ao método da REML na estimação de σ_p^2 e apresentou menos desvantagens do que o método da REML na estimação de σ_{pa}^2 e σ_{pta}^2 . O maior problema está na predição de σ_{pt}^2 , porém esse componente só foi estimado satisfatoriamente pelo método da REML no caso de um alto valor de referência, nos

outros casos o erro apresentado pela REML foi quase tão alto quanto o da ANOVA.

5.2 Parâmetros Genéticos

O método da ANOVA se mostrou: pouco satisfatório, porém mais eficiente, ou no mínimo tão eficiente quanto o método da REML na predição das interações entre as variâncias aditiva e de ambiente (V_{ae}), de dominância e de ambiente (V_{de}) e genotípica e de ambiente (V_{ge}); viesado porém mais eficiente do que o método da REML na predição do grau médio de dominância GMD ; e muito satisfatório e mais eficiente do que o método da REML na predição das variâncias aditiva (V_a) e fenotípica (V_f) e na predição do coeficiente de herdabilidade (h^2), com exceção para o caráter número de folhas, para o qual o método da ANOVA foi pouco satisfatório, porém mais eficiente que o da REML na predição de h^2 ; e o método da REML foi pouco satisfatório, porém mais eficiente que o da ANOVA na predição de V_f .

O método da REML se mostrou pouco satisfatório, porém mais eficiente, ou pelo menos tão eficiente quanto o método da ANOVA somente na predição da variância de dominância (V_d).

Pode-se concluir que, do ponto de vista genético, o método da ANOVA é bem mais eficiente do que o método da REML na predição das variâncias genéticas, coeficiente de herdabilidade e grau médio de dominância; fornecendo estimativas muito satisfatórias ou satisfatórias para a maioria dos parâmetros. As maiores desvantagens ocorreram para o caráter número de folhas, aquele que apresentou menores valores de referência para os componentes de variância do modelo matemático 6, variâncias genéticas e grau médio de dominância apesar de apresentar um valor alto para o coeficiente de herdabilidade. Ainda assim o método da ANOVA apresentou uma menor taxa de erros do que o método da REML.

Deve se tomar muito cuidado, no entanto, com a estimativa do grau médio de dominância, pois para todos os caracteres o método da ANOVA, apesar de ser o mais eficiente apresentou viés na predição, superestimando em cerca de 45% o valor que deveria estimar. Tal resultado se deve ao fato do método da ANOVA ter sido pouco eficiente na estimação da variância de dominância, fornecendo estimativas superestimadas em cerca de 100% do valor que deveria estimar para essa variância. Esse resultado é bastante surpreendente e

preocupante, já que o delineamento III foi desenvolvido, inicialmente, com o propósito de fornecer uma estimativa mais confiável para o grau médio de dominância.

REFERÊNCIAS

ALLARD, R.W. **Princípio do melhoramento genético das plantas**. Rio de Janeiro: USAID, 1971. 381p.

ARIAS, C.A.A.; SOUZA JÚNIOR, C.L. Genetic variance and covariance components related to intra- and interpopulation recurrent selection in maize (*Zea mays* L.). **Genetics and molecular biology**, São Paulo, v.21, n.4, p.537-544, Dec. 1998.

BARBIN, D. **Componentes de variância: teoria e aplicações**. Piracicaba: FEALQ, 1993. 120p.

BOCA, R.T.; CANTET, R.J.C. On the precision of REML estimates of additive and dominance variance components in maize using relationship matrices. **Maydica**, Roma, v.49, p.295-302, 2004

CARNEIRO JÚNIOR, J.M. et al. Avaliação de métodos de estimação de componentes de variância utilizando dados simulados. **Revista Brasileira de Zootecnia**, Viçosa, v.33, n.2, p.328-336, mar./abr. 2004.

CECON, P.R. **Alternativas de análise de experimentos em látice e aplicações no melhoramento vegetal**. 1992. 109p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1992.

COCHRAN, W.G.; COX, G.M. **Experimental designs**. 2.ed. New York: John Wiley & Sons, 1957. 611p.

COMSTOCK, R.E.; ROBINSON, H.F. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. **Biometrics**, Oxford, v.4, n.4, p.254-266, Dec. 1948.

COMSTOCK, R.E.; ROBINSON, H.F. Estimation of average dominance of genes. In: GOWEN, J.W. (Ed.) **Heterosis**. Ames: Iowa State College Press, 1952. cap 30, p.494-516.

CUSTÓDIO, T.N. **Comparação de modelos mistos para dados de proporção visando a estimação do coeficiente de herdabilidade**. 2004. 121p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2004.

EMBRAPA. Disponível em:
<<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Milho/CultivodoMilho/importancia.htm>>. Acesso em: 6 jan. 2009

FALCONER, D.S. **Introduction to quantitative genetics**. 3.ed. London: Longman Group, 1993. 438p.

FERNANDEZ, D.W.X. **Modelos de populações finitas e máxima verossimilhança restrita no problema de estimativas negativas para componentes de variância.** 1991. 118p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1991.

FREITAS, A.R. **Comparação de métodos para estimação de componentes de variância e parâmetros afins de múltiplos caracteres em bovinos.** 1991. 170p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1991.

GOMES, F.P. **Curso de estatística experimental: II parte - experimentos complexos.** Piracicaba: ESALQ, 1958.

KEARSEY, M.J.; POONI, H.S. **The genetical analysis of quantitative traits.** Cheltenham: Stanley Thornes, 1998.

LITTELL, R.C. et al. **SAS for mixed models.** 2.ed. Cary: SAS Institute, 2006. 813p.

MOLL, R.H.; LINDSEY, M.F.; ROBINSON, H.F. Estimates of genetic variances and level of dominance in maize. **Genetics.** Pittsburgh, v.49, p.411-423, Mar. 1964

OTSUK, I.P. **Influência de transformações de dados sobre a estimação dos componentes de variância.** 1991. 66p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1991.

RAMALHO, M.A.P. **Eficiência relativa de alguns processos de seleção intrapopulacional do milho baseados em famílias não endógamas.** 1977. 122p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1977.

RAO, P.S.R.S. **Variance components estimation: mixed model, methodologies and applications.** New York: Chapman & Hall, 1999.

REGAZZI, A.J. et al. Análise de experimentos em látice quadrado com ênfase em componentes de variância: Análise conjunta. **Pesquisa Agropecuária Brasileira,** Brasília, v.34, p.1987-1997, nov. 1999.

RENCHE, A.C. **Linear models in statistics.** New York: John Wiley & Sons, 2000. 578p.

SEARLE, S.R.; CASELLA, G.; McCULLOCH, C.E. **Variance components.** New York: John Wiley, 1992. 501p.

SILVA, A.R.da **Análise da caracteres quantitativos em milho com o delineamento III e marcadores moleculares.** 2002. 143p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2002.

SILVA, A.R.da et al. Estimates of genetic variance and level of dominance in a tropical maize population. I. grain yield and plant traits. **Maydica,** Roma, v.49, p.65-71, 2004.

SILVA, H.D. **Análise de experimentos em látice quadrado ("square lattice" com ênfase em componentes de variância e aplicações no melhoramento genético vegetal**. 1997. 139p. Dissertação (Mestrado em Gentética e Melhoramento) - Universidade Federal de Viçosa, Viçosa, 1997.

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

VILARINHO, A.A. **A importância do melhoramento genético na cultura do milho**. Disponível em: <<http://www.universia.com.br/materia/materia.jsp?id=8315>>. Acesso em: 18 dez. 2008

WARDYN, B.M.; EDWARDS, J.W.; LAMKEY, K.R. The genetic structure of a maize population: the role of dominance. **Crop Science**, Madison, v.47, p.467-476, Mar./Apr. 2007.

WOLF, D.P.; PETERNELLI, L.A.; HALLAUER, A.R. Estimates of genetic variance in an F_2 maize population. **American Genetic Association**, Newport, v.91, p.384-391, 2000.