



VINÍCIUS TEIXEIRA ANDRADE

**IMPLICAÇÕES DA MODELAGEM
ESTATÍSTICA NA SELEÇÃO DE PROGÊNIES
DE CAFEIEIRO**

LAVRAS - MG

2013

VINÍCIUS TEIXEIRA ANDRADE

**IMPLICAÇÕES DA MODELAGEM ESTATÍSTICA NA SELEÇÃO DE
PROGÊNIES DE CAFEIEIRO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas para a obtenção do título de Mestre.

Orientadora

Dra. Flávia Maria Avelar Gonçalves

LAVRAS - MG

2013

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Andrade, Vinícius Teixeira.

Implicações da modelagem estatística na seleção de progênies
de cafeeiro / Vinícius Teixeira Andrade. – Lavras : UFLA, 2013.

94 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2013.

Orientador: Flávia Maria Avelar Gonçalves.

Bibliografia.

1. Melhoramento genético de plantas. 2. Medidas repetidas. 3.
Produção de grãos. 4. Valor reprodutivo. I. Universidade Federal de
Lavras. II. Título.

CDD – 631.523

VINÍCIUS TEIXEIRA ANDRADE

**IMPLICAÇÕES DA MODELAGEM ESTATÍSTICA NA SELEÇÃO DE
PROGÊNIES DE CAFEIEIRO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas para a obtenção do título de Mestre.

APROVADA em 29 de abril de 2013.

Dr. José Airton Rodrigues Nunes

UFLA

Dr. César Elias Botelho

EPAMIG

Dra. Flávia Maria Avelar Gonçalves
Orientadora

LAVRAS-MG

2013

Aos meus pais por me darem a oportunidade da vida e que, por algumas vezes, não me permitiram interromper os estudos para seguir meus desejos momentâneos. Vocês conseguem aumentar seus valores reprodutivos a cada dia. Antes do estudo já gostava da Agricultura e da Natureza, hoje a afinidade é ainda maior.

Aos meus irmãos, com os quais tenho mais que uma relação de parentesco, e sim uma profunda amizade. Sentimento que se estende aos familiares e amigos. Familiares que atualmente se expandiram com a chegada da Ranízia e da Antonella (sobrinha e afilhada) e podem se estender ainda mais com a presença da Flávia em minha vida.

Aos professores práticos e teóricos, que perpetuam o conhecimento para as futuras gerações. À todas as pessoas com quem convivi e com as quais troquei alguma experiência.

Aos homens e mulheres do “campo”, com os quais compartilho a mesma paixão.

DEDICO

AGRADECIMENTOS

Meus agradecimentos dirigem-se mais às instituições que fizeram parte de minha formação. Como instituições são feitas de pessoas não estarei sendo negligente. Em primeiro lugar agradeço à fazenda de meu pai. Lugar de inspiração e que me despertou o gosto pelo caminho que sigo. Aos amigos de Conceição da Barra de Minas (Cassiterita), que me ensinaram mais do que se encontra nos livros. À UFLA, realizadora de sonhos. Ao Rehagro, local de muito aprendizado e que me permitiu conviver em algumas fazendas. Dentre as quais cito a Fazenda Samambaia, Dona Neném e São João Grande. Atualmente, não posso deixar de agradecer às Fazendas Boa Vista e Lagoa onde passo momentos agradáveis durante o trabalho.

Agradecimento especial à Empresa de Pesquisa Agropecuária de Minas Gerais-EPAMIG. Obrigado pela oportunidade de convivência e por ter me iniciado na pesquisa. Agradeço também aos funcionários da Fazenda Experimental de Machado, onde os dados foram cuidadosamente obtidos durante mais de uma década. Ao Consórcio de Pesquisa Café e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq, pela concessão das bolsas de estudo desde o fim da graduação. E ao Instituto Nacional de Ciência e Tecnologia- INCT-Café, pelo apoio na condução das pesquisas.

E por último, gostaria de agradecer ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas e ao Departamento de Biologia, por ter me permitido viver nesse ambiente. Obrigado pelo prazer da convivência e ensinamentos transmitidos, os quais transcendem à Genética e Melhoramento de Plantas.

“Não há boa prática sem uma correta teoria”

CHARLES DARWIN

RESUMO

Para se obter ganhos em produtividade é necessário que as progênies selecionadas sejam realmente superiores geneticamente. Para se alcançar tal objetivo, os requisitos indispensáveis se constituem numa fenotipagem fidedigna, mediante uma experimentação adequada e a subsequente exploração minuciosa dos dados experimentais por meio da adoção de métodos estatísticos acurados. Os dados de produção de grãos do cafeeiro são oriundos de colheitas anuais, que são medições repetidas nas mesmas parcelas experimentais, gerando dados longitudinais no tempo. A característica principal nesse tipo de análise é a covariância existente nos fatores aleatórios do modelo estatístico. Fato que nem sempre é levado em conta corretamente pelos modelos utilizados. Objetivou-se neste trabalho avaliar diferentes estruturas de matrizes de covariância na modelagem do comportamento produtivo de progênies de cafeeiro e suas implicações na seleção e verificar a influência do modo como se consideram os dados ao analisá-los. Foram analisadas 21 progênies $S_{0,1}$ de *Coffea arabica* durante oito colheitas. Como estratégias de agrupamento de dados analisaram-se as colheitas anuais, em biênios, focando somente colheitas altas e também os anos de baixa produção. Modelou-se, para cada situação, a matriz de covariância referente ao resíduo (R) e à interação progênies com colheitas (PC). Detectaram-se, por meio de BIC, abordagens mais condizentes com o caráter. As consequências observadas foram a alteração na estimativa de parâmetros populacionais, no ordenamento das progênies e também no índice de coincidência ao se selecionar as melhores. Portanto, diferentes estruturas das matrizes de covariância devem ser analisadas em todas as situações, quando forem requisitadas pelos dados. As produções não devem ser analisadas em biênios e sim por meio de suas colheitas anuais ou, caso necessário, apenas pelos anos de alta produção. A não observância desses fatores pode induzir à recombinação de progênies tidas como superiores quando na verdade não o são. O resultado desse processo no ganho realizado com a seleção não é desejável no melhoramento de qualquer espécie vegetal.

Palavras-chave: Melhoramento Genético de Plantas. Medidas repetidas. Produção de Grãos. Valor reprodutivo.

ABSTRACT

In order to obtain gains in productivity it is necessary that the progenies selected be genetically superior. To achieve such an objective, the essential requisites are a reliable phenotyping done through an adequate experimentation and the subsequent meticulous exploration of the experimental data by adopting accurate statistical methods. The data from coffee grain production derive from harvests in each year (repeated measurements of the same experimental plots, generating longitudinal data in time). The main characteristic of this type of analysis is the covariance existent in the random factors of the statistical model. Thus, the objectives of this work were to evaluate different covariance matrix structures in modeling the productive behavior of coffee progenies and its implications in selection, and verifying the influence of the manner in which the data are considered in analysis. We analyzed 21 $S_{0.1}$ *Coffea arabica* progenies during eight harvests. We analyzed the annual harvests in biennium, only large harvests and years in which the harvests were low, as a strategy in considering the data. We modeled for each situation, the covariance matrix referent to residue (R) and to the interactions between progenies and harvest (PH). We detected by BIC, approaches that promote better fit to the character. The consequences observed were the alterations in the estimation of population parameters, in progeny rank order and also in the coincidence index in selecting the best. Therefore, different covariance matrix structures must be analyzed in all situations when requested by the data. The productions must not be analyzed in biennium, but in by means of annual harvests or, if necessary, only in the years with large productions. The non-observation of these factors may induce the recombination of progenies considered superior when, in reality, they aren't. The result of this process in the gain obtained with selection isn't desirable in the genetic improvement of any plant species.

Keywords: Plant breeding. Repeated measures. Bean yield. Reproductive value.

LISTA DE TABELAS

Tabela 1	Relação de progênies $S_{0:1}$ provenientes dos cruzamentos entre linhagens de Mundo Novo com Mundo Novo e Mundo Novo com Bourbon Amarelo avaliadas em Machado, Minas Gerais.....	35
Tabela 2	Estimativa da média, do coeficiente de variação ambiental e da herdabilidade na média das progênies para colheitas individuais e em biênios.....	52
Tabela 3	Correlações de Spearman dos E-BLUP's de progênies entre as colheitas anuais.....	55
Tabela 4	Correlações de Spearman dos E-BLUP's de progênies entre os biênios.....	55
Tabela 5	Valores P do teste de esfericidade de Mauchly para as colheitas individuais, biênios, anos de alta (bons) e baixa (ruins) produtividade.....	56
Tabela 6	Probabilidade (\hat{W}_i) , em porcentagem, de o modelo da estrutura de covariância residual (R) ser o correto nos diferentes modos de se considerar os dados (anual, biênio, anos bons e anos ruins).....	58
Tabela 7	Probabilidade (\hat{W}_i) , em porcentagem, de o modelo da estrutura de covariância da interação progênies com colheitas (PC) ser o correto a partir da melhor estrutura residual escolhida na Tabela 6, nos diferentes modos de se considerar os dados.....	59
Tabela 8	Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênies das modelagens escolhidas para colheitas anuais.....	60

Tabela 9	Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênes das modelagens escolhidas para biênios	61
Tabela 10	Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênes das modelagens escolhidas para os anos bons.....	62
Tabela 11	Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênes das modelagens escolhidas para os anos ruins	63
Tabela 12	Correlação de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas anuais	64
Tabela 13	Índice de coincidência de Hamblin e Zimmerman (%) na seleção das cinco melhores progênes por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas anuais	65
Tabela 14	Correlações de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas agrupadas em biênios	66
Tabela 15	Índice de coincidência de Hamblin e Zimmerman (%) na seleção das cinco melhores progênes por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas agrupadas em biênios	67
Tabela 16	Correlação de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra alta (Anos Bons)	67
Tabela 17	Índice de coincidência de Hamblin e Zimmerman (%) na seleção das cinco melhores progênes por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra alta (Anos Bons)	68
Tabela 18	Correlação de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra baixa (Anos Ruins).....	69

Tabela 19 Índice de coincidência de Hamblim e Zimmerman (%) na seleção das cinco melhores progênies por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra baixa (Anos Ruins)	69
Tabela 20 Correlação de Spearman entre os modos de se agrupar os dados para seleção.....	70
Tabela 21 Índice de coincidência (%) na seleção das cinco melhores progênies em cada modo de se agrupar os dados para seleção	70

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	15
2.1	Melhoramento genético do café arábica	15
2.1.1	Melhoramento da cultivar mundo novo	18
2.2	Aspectos estatísticos do melhoramento do cafeeiro	20
2.2.1	Modelagem estatística	22
2.2.2	Análise de dados longitudinais	27
3	MATERIAL E MÉTODOS	34
3.1	Análise dos dados	36
3.2	Estruturas de covariância	40
3.2.1	Simetria composta (CS)	41
3.2.2	Antedependência (ANTE)	41
3.2.3	Autorregressiva (AR)	42
3.2.4	Autorregressiva com variâncias heterogêneas (ARH)	42
3.2.5	Médias móveis autorregressiva (ARMA)	43
3.2.6	Simetria composta com variâncias heterogêneas (CSH)	43
3.2.7	Fator Analítica (FA)	45
3.2.8	Huynh-Feldt (HF)	46
3.2.9	Toeplitz (TOEP)	46
3.2.10	Toeplitz com Variâncias Heterogêneas (TOEPH)	47
3.2.11	Não estruturada (UN)	47
3.3	Estimação de parâmetros	48
3.4	Correlação de Spearman e índice de coincidência	49
4	RESULTADOS	51
4.1	Análise exploratória	51
4.2	Estruturas de covariância	57
4.3	Consequências na seleção	59
4.3.1	Estimativa dos parâmetros	60
4.3.2	Alteração no ordenamento	64
5	DISCUSSÃO	71
6	CONCLUSÃO	82
	REFERÊNCIAS	83
	APÊNDICE	93

1 INTRODUÇÃO

O melhoramento genético de características quantitativas como a produção de grãos do cafeeiro é altamente influenciado pelo ambiente. Com isso a identificação de indivíduos superiores quanto ao valor reprodutivo é dificultada. Esse fato afeta a seleção, acarretando ganhos genéticos pequenos, ou até mesmo nulos. O impacto dessa ocorrência agrava-se quando se trabalha com plantas perenes. Principalmente em função de o processo ser longo e dispendioso. Corre-se o risco de, no final do ciclo de melhoramento, ou no seu decorrer, que os genótipos obtidos não superem o desempenho das cultivares em uso. Na cafeicultura esse desafio é enorme tendo-se em vista os ganhos que foram obtidos com a cultivar Mundo Novo há mais de meio século.

Para se obter ganhos produtividade de grãos é necessário que as progênies selecionadas sejam realmente superiores geneticamente. Para se alcançar tal objetivo, os requisitos indispensáveis se constituem numa fenotipagem fidedigna, mediante uma experimentação adequada e a subsequente exploração minuciosa dos dados experimentais por meio da adoção de métodos estatísticos acurados. No melhoramento de plantas perenes essa tarefa merece notável atenção devido à tomada de várias mensurações na mesma parcela, a exemplo das colheitas. Esse procedimento gera dados do tipo longitudinais no tempo. Suas principais características estão relacionadas com a estrutura das correlações existentes entre essas medições e a presença ou não de heterogeneidade de variâncias. Esses aspectos elevam a complexidade dos modelos estatísticos, e tendem a provocar distorções na seleção dos indivíduos superiores quando se utiliza modelos estatísticos não condizentes com a natureza biológica do caráter, e/ou métodos pouco robustos para estimação de parâmetros.

Dentre os métodos utilizados para a análise de medidas repetidas destacam-se a análise de variância com medidas repetidas, a análise multivariada e a modelagem das médias e estruturas de covariância dos fatores do modelo estatístico. A primeira faz suposições mais restritivas quanto à estrutura de covariâncias dos dados, podendo afetar as inferências em muitas situações. A técnica multivariada é uma boa opção. No entanto, o excesso de parâmetros a serem estimados pode tornar a análise menos robusta. A última abordagem é comumente citada na literatura como modelagem mista. Atualmente, a modelagem mista é considerada ideal em ciências sociais, econômicas e biológicas. Em função das características do melhoramento do cafeeiro essa técnica mostra-se promissora.

O melhoramento para produção no cafeeiro envolve a avaliação da produção dos genótipos por aproximadamente quatro colheitas em cada geração. Atualmente, esses dados são tratados de uma forma que pode não ser condizente com a realidade da cultura, que apresenta algumas peculiaridades, como a bienalidade. Nos modelos estatísticos comumente utilizados a correlação entre as colheitas é considerada constante, bem como suas variâncias. A adoção desses procedimentos pode afetar a predição do valor genético do indivíduo ou de sua progênie. Não considerar modelagens estatísticas específicas, além de promover alteração no valor estimado dos parâmetros, pode até mesmo provocar alteração no ordenamento das progênies avaliadas. Alguns trabalhos detectaram a alteração na estimação de parâmetros em função das estruturas de covariância utilizadas (APIOLAZA; GILMOUR; GARRICK, 2000). No entanto a quantificação do impacto no processo seletivo não foi encontrada.

Portanto, objetivou-se neste trabalho avaliar diferentes estruturas de matrizes de covariância na modelagem do comportamento produtivo de progênies de cafeeiro e suas implicações na seleção e verificar a influência do modo como se consideram os dados ao analisá-los.

2 REFERENCIAL TEÓRICO

2.1 Melhoramento genético do café arábica

O programa de melhoramento genético do cafeeiro no Brasil iniciou-se em 1933, no Instituto Agrônomo de Campinas (IAC), com a criação das seções de genética e de café. O plano de trabalhos constituía-se basicamente dos seguintes setores: seleção de cafeeiros matrizes de cultivares comerciais e isolamento de progênies; geração de novos genótipos por meio da hibridação intra e interespecífica; introdução e estudo de novas cultivares comerciais; pesquisa sobre novos genótipos que ocorrem em plantações como consequência de hibridações e/ou mutações (CARVALHO; FAZUOLI, 1993). Posteriormente, outros programas foram criados em outras instituições, com destaque para o convênio EPAMIG/UFLA/UFV e o MAPA/PROCAFÉ em Minas Gerais, o programa do IAPAR, no Paraná e o INCAPER no Espírito Santo (MEDINA FILHO; BORDIGNON; CARVALHO, 2008). Os trabalhos realizados proporcionaram grandes avanços no melhoramento genético do cafeeiro no Brasil, disponibilizando, para o agricultor, cultivares com elevado potencial produtivo, vigor, rusticidade, uniformidade de maturação e resistência à ferrugem, além de outras características de interesse agrônomo e comercial (AGWANDA et al., 2003).

A espécie *Coffea arabica* é alotetraplóide segmental, com $2n = 44$ cromossomos, autocompatível e multiplica-se, predominantemente, por autofecundação, sendo a taxa de fecundação cruzada em torno de 10% (CARVALHO; KRUG, 1949). Devido a essa autogamia predominante e o fato de ser propagado comercialmente por sementes, o melhoramento genético visa obtenção de cultivares homozigóticas.

A heterose para produção foi relatada por alguns trabalhos. O vigor híbrido pode atingir valores de 10 a 200% em relação ao melhor genitor (WALYARO, 1983). Para verificar o progresso genético com o uso de híbridos, pesquisadores da Costa Rica cruzaram variedades selvagens do Sudão e Etiópia com cultivares americanas. Os resultados apontam superioridade produtiva dos híbridos em relação à média dos genitores, na ordem de 34 a 58% (BERTRAND et al., 2011). No entanto, os genitores utilizados possuem média baixa, o que deve ter conduzido a resultados tão expressivos. Essa constatação também é observada no trabalho de Cilas et al. (1998), no qual os autores realizaram um dialelo parcial. Além disso, a propagação do híbrido para os produtores constitui um problema. A obtenção de sementes híbridas por meio do cruzamento entre linhagens torna-se restritiva pelo reduzido número de sementes por cruzamento. Técnicas de clonagem em *Coffea arabica* estão sendo desenvolvidas com sucesso, porém o custo da muda obtida ainda é alto (CARVALHO et al., 2008; ETIENNE et al., 2002).

O melhoramento do cafeeiro inicia-se pela obtenção da população segregante gerada por delineamentos de cruzamento. Após a obtenção dessa população ela pode ser conduzida por diversos métodos, podendo-se destacar bulk, bulk dentro de progênie e o genealógico, sendo esse último o mais utilizado para a condução da população segregante (MEDINA FILHO; BORDIGNON; CARVALHO, 2008; SAKIAMA et al., 2005). Esse método não apresenta vantagem em relação aos demais. Exceto se essa genealogia for utilizada na predição do valor genético dos indivíduos (RAMALHO; CARVALHO; NUNES, 2013).

A escolha das estratégias adotadas para se melhorar o cafeeiro são função da herdabilidade do caráter, dos recursos disponíveis e da habilidade do melhorista (RAMALHO et al., 2012). A condução da população é realizada até se atingir a homozigose necessária para se selecionar os melhores indivíduos. O

longo prazo requerido no processo seletivo é função da necessidade de se avaliar os genótipos no decorrer do tempo e ao período juvenil do cafeeiro. Há estimativas de que a liberação de uma nova cultivar de café necessite de aproximadamente quinze a vinte anos. No entanto, estratégias vêm sendo desenvolvidas para se reduzir o tempo do ciclo de melhoramento (RAMALHO; CARVALHO; NUNES, 2013; SERA, 2001).

Como o foco do melhoramento é no ganho com a seleção em um ciclo, toda prática que vise diminuir o intervalo entre gerações é benéfico. A acumulação de alelos favoráveis é um processo gradativo conseguido por meio da seleção e recombinação contínua entre os melhores indivíduos. Esse fato é tão mais importante quanto maior for a influência ambiental sobre o caráter em estudo (BERNARDO, 2010). Para prever o ganho com a seleção para produtividade de grãos em progênies de café Arábica com alelos de resistência à ferrugem, Oliveira et al. (2011) encontraram resultados positivos. Foram avaliadas noventa e oito progênies em quatro colheitas, utilizando quatro modos de se considerar os dados para se realizar a seleção. Analisaram-se a média das quatro colheitas, colheitas anuais, média dos melhores anos e média dos anos com produção mais baixa. O interessante nos resultados obtidos é que o maior ganho em porcentagem em relação à média original foi para as colheitas com baixa produção. Na média das duas safras baixas o valor foi de 103,30%. O pior resultado foi obtido utilizando-se a média das produções, cujo valor foi de 20,8%. Apesar da maior porcentagem de ganho com a seleção obtida nas colheitas de baixa produção, selecionar com base nesses dados não foi recomendado pelos autores, uma vez que apresentaram baixas correlações fenotípicas e genéticas com a média das quatro colheitas. Os autores argumentam que a seleção com base na primeira alta safra é mais indicada para o melhoramento do cafeeiro. Esses resultados condizem com o relatado por Sera

(1987) que foi um dos primeiros pesquisadores a estudar a viabilidade da seleção precoce no cafeeiro.

2.1.1 Melhoramento da cultivar mundo novo

Dentre as cultivares disponíveis para o plantio no Brasil a ‘Mundo Novo’ ocupa lugar de destaque, apresentando boas produções em quase todas as regiões cafeeiras do Brasil (FAZUOLI et al., 2005). As primeiras seleções que deram origem à cultivar foram realizadas em 1943. Pesquisadores do IAC foram informados de uma lavoura produtiva e rústica da cultivar Sumatra na Fazenda Aparecida, no município de ‘Mundo Novo’, hoje Urupês, na região araraquarense do estado paulista. A lavoura continha quatorze mil cafeeiros com doze anos de idade, os quais apresentavam variação quanto à sua conformação. Em conversa com produtores da época descobriu-se que a lavoura da fazenda Aparecida tinha sido formada por sementes provindas de uma plantação em Jaú, em 1928. Rastreamento de cafeeiros com características favoráveis foram coletados sementes e propágulos para a enxertia em mais nove ocasiões até 1949. Essas coletas foram realizadas nas lavouras mais antigas da região buscando-se a origem da Mundo Novo. Relatos indicaram que essas plantações provinham de uma só planta considerada como Sumatra, situada entre os municípios de Mineiros do Tietê e Barra Bonita (CARVALHO et al., 1952).

Progenies das matrizes aí selecionadas foram plantadas nas estações experimentais do Instituto Agrônomo de Campinas, em Ribeirão Preto, Pindorama, Mococa e Jaú. As análises dos primeiros seis anos de produção dessas progenies revelaram tratar-se de genótipos promissores em todas as cinco localidades estudadas. As progenies apresentaram elevada capacidade e estabilidade produtiva, associada à boa rusticidade. Embora algumas progenies expressassem características indesejáveis, como a ocorrência de plantas

improdutivas e a produção de frutos com alta incidência de lojas vazias, sem sementes normais.

A cultivar Mundo Novo foi resultado do cruzamento natural entre as variedades 'Bourbon Vermelho' e 'Sumatra' (variedade Typica), sendo essa vigorosa, produtiva e de porte alto, apresentando internódios longos, cujas plantas, quando adultas, atingem de 3 a 4 metros de altura. Os pesquisadores da época notaram grande variabilidade fenotípica, visto tratar-se de um material heterogêneo, resultante de hibridação entre variedades distintas. Em experimentos regionais em Campinas, São Paulo, verificou-se que as melhores progênies de 'Mundo Novo' chegaram a produzir 80% mais que as melhores seleções de 'Bourbon Amarelo', 95% mais que as melhores de 'Bourbon Vermelho' e 240% mais que as progênies da variedade 'Typica', destacando-se as progênies LCMP-376-4 e CP-379-19 (CARVALHO et al., 1952, 1961).

Atualmente, as progênies selecionadas de 'Mundo Novo' vêm sendo utilizadas em cruzamentos intraespecíficos e também em hibridações interespecíficas para formação de populações de melhoramento. Das hibridações já conseguidas, de interesse econômico, destacam-se as cultivares Catuaí Amarelo, Catuaí Vermelho, Icatu, Rubi, Topázio, Ouro Verde e Travessia. A grande adaptação das progênies de 'Mundo Novo', nas mais diversas condições de ambiente e a sua boa capacidade de combinação nas hibridações evidenciam o interesse da pesquisa para a obtenção de novas progênies dessa cultivar (CARVALHO et al., 2006).

Em várias pesquisas tem-se estudado o comportamento de progênies de 'Mundo Novo', em diferentes regiões do Brasil. Fazuoli et al. (2000) verificaram que as melhores progênies foram IAC 376-4, IAC 379-19, IAC 382-14, IAC 388-17, IAC 515-11, IAC 464-12, IAC 467-11, IAC 502 e IAC 480-6. Rocha, Carvalho e Fazuoli (1980) indicaram, para a região de Mococa, as seguintes progênies como mais promissoras: IAC 388-17-1, IAC 472, IAC 388-6-17, IAC

440-14 e IAC 505-6. Almeida e Carvalho (1986) relataram que as progênies IAC 376-4, IAC 388-17-1, IAC 515-11, IAC 388-6-17, IAC 515-3 e IAC 376-4-22 revelaram-se melhores para a região de Varginha, em Minas Gerais. Begazo, Vieira e Paula (1977) concluíram que as linhagens IAC 500-1, IAC 464-15 e IAC 480 foram mais produtivas em Viçosa, também Minas Gerais. Por aproximadamente 33 anos estudou-se a produção de progênies de ‘Mundo Novo’ e verificou-se que algumas progênies apresentaram longevidade e se destacaram como as mais produtivas no período de 1947 a 1979. Fazuoli et al. (2005) discutem, à luz desses resultados, que as melhores progênies devem possuir capacidade produtiva, longevidade e rusticidade.

Em experimentos da Seção de Genética do IAC, as linhagens de ‘Mundo Novo’ apresentaram uma produção média anual de 2000 Kg/ha, inclusive nas primeiras produções após o plantio, com uma oscilação variando de 1500 a 3000 Kg/ha. Em anos de elevada produção, pode-se atingir até 6.000 Kg/ha de café beneficiado (ROCHA; CARVALHO; FAZUOLI, 1980). Posteriormente, Carvalho et al. (2006) encontraram resultados semelhantes trabalhando com progênies de Mundo Novo em Machado, sul de Minas Gerais, em experimento instalado na fazenda experimental da EPAMIG e concluíram que as melhores progênies, ao longo das quatorze colheitas analisadas, foram as de melhor desempenho nas primeiras produções.

2.2 Aspectos estatísticos do melhoramento do cafeeiro

Uma das práticas comuns no melhoramento de plantas, como em praticamente todas as Ciências Biológicas é se fazer comparações entre os indivíduos a partir de dados fenotípicos. No melhoramento essa comparação é realizada visando-se à seleção. A seleção é a separação em categorias dos constituintes de uma população. O processo consiste em selecionar indivíduos

avaliados que apresentem os melhores valores reprodutivos preditos (HEDRICK, 2005; LEHMANN, 1961; LUSH, 1947; NAKAGAWA; SCHIELZETH, 2010).

No melhoramento de plantas, a eficiência do processo seletivo depende de vários fatores. Um deles refere-se à fenotipagem fidedigna dos caracteres alvo de melhoramento dos genótipos sob teste. Isso está diretamente vinculado à adoção de delineamentos experimentais apropriados e à observância de aspectos relacionados ao dimensionamento dos experimentos (número de repetições, tamanho de parcela, etc.). Outro ponto importante para melhoria da eficiência da seleção é o emprego de métodos estatísticos que possibilitem explorar ao máximo os dados experimentais e informações adicionais acerca dos genótipos (e.g. grau de parentesco existente, genotipagem, etc.), resultando em estimativas ou predições acuradas dos valores genéticos ou *breeding values* (RESENDE, 2007; RESENDE; DUARTE, 2007).

Alguns aspectos biológicos do cafeeiro, como o ciclo reprodutivo longo, que necessita de medidas repetidas, a acentuada oscilação anual de produção (bienalidade), a sobreposição de gerações, entre outros, conduzem a algumas peculiaridades em termos de melhoramento. Outra questão é a redução da taxa de sobrevivência dos experimentos durante sua vida útil, fato que gera dados desbalanceados. O desbalanceamento pode também ocorrer porque algumas plantas não produzem em determinada medição/colheita. Para o cafeeiro também é relevante a seleção em nível de planta, e não somente com base na média de progênie. Em função do exposto, tem-se a necessidade do uso de métodos especiais de estimação de parâmetros genéticos e predição de valores genotípicos (PETEK; SERA; FONSECA, 2008; RESENDE, 2007; RESENDE; DIAS, 2000; RESENDE et al., 2001).

2.2.1 Modelagem estatística

A principal função dos modelos estatísticos é procurar descrever a realidade de algum fenômeno natural. No caso de experimentos, por se constituírem em inquisições planejadas, tem-se a resposta (fenótipos observados) em função de fatores previamente escolhidos pelo pesquisador e de suas interrelações (efeitos de interação ou hierarquizados). Fatores principais no melhoramento de plantas referem-se aos genótipos, colheitas e ambientes.

Outro aspecto importante na modelagem é a especificação da natureza dos efeitos presentes no modelo estatístico, o qual pode ser fixo ou aleatório. A definição de como o efeito será considerado, em certas ocasiões, pode não ser tarefa fácil. No entanto, mais importante do que aspectos teóricos, deve-se ponderar o contexto da pesquisa, pois em se considerando um efeito como aleatório, a inferência é ampliada para a população de referência, enquanto que para efeitos fixos, a inferência é restrita aos níveis do fator em estudo. Resende (2007) afirma que para fins de melhoramento genético, efeitos de genótipos devem ser considerados como aleatórios. O autor argumenta que a seleção durante as etapas de avaliação e, principalmente, a recomendação de cultivares aos agricultores foca o desempenho futuro das plantas. Ou seja, como esse genótipo se comportará quando for cultivado. Contrariamente, os fatores referentes aos efeitos ambientais devem ser preferencialmente considerados como fixos. Como o interesse reside no efeito genético, caso o fator de blocagem seja considerado aleatório, por exemplo, esse terá seu efeito penalizado pelo regressor. Consequentemente se adicionarão ruídos à estimativa do valor genético. Além disso, os efeitos de blocos, colheitas e ambientes dificilmente se repetirão em ocasiões futuras. Dessa forma, tais fatores devem

ser considerados aleatórios desde que tenham pelo menos dez níveis. E também, quando os blocos forem incompletos, pois existe informação disponível a ser recuperada entre blocos dentro da repetição (PIEPHO; BUSCHE; EMRICH, 2003; RESENDE, 2007; SMITH; CULLIS; THOMPSON, 2005). Vários autores têm apresentado uma discussão conceitual sobre essa questão (BEARZOTI, 2002; EISENHART, 1947; SEARLE; HENDERSON, 1991).

Existem três tipos de modelagem com foco na seleção. A primeira, denominada tipo I ou modelo fixo, considera todos os fatores de efeito fixo, exceto o erro. O tipo II, modelo aleatório, apenas a constante populacional é de efeito fixo. Já a modelagem do tipo III é mista, coexistindo fatores de efeito fixo e aleatório, além da constante e do erro aleatório (EISENHART, 1947; HENDERSON, 1975).

Um aspecto fundamental para fins de melhoramento, independente do modelo empregado, é a busca por estimadores ou preditores que resultem em estimativas ou predições mais próximas possíveis dos verdadeiros *breeding values* dos genótipos em teste (WHITE; HODGE, 1988). Para a seleção sob modelo fixo, temos as médias fenotípicas ajustadas dos genótipos, que se constituem nos melhores estimadores lineares não viesados dos *breeding values* (BLUE). Nesse caso, aplicam-se os métodos estatísticos dos quadrados mínimos, que se constitui na base para o método da análise de variância. Para a seleção sob modelo aleatório ou misto tem-se o método geral proposto por Henderson (1950), no qual a seleção dos genótipos seria pautada com base no preditor BLUP, que também se constitui numa função linear dos fenótipos, de elevada acurácia e não viciado. Mas, para que o BLUP se constitua absolutamente no mais eficiente dentre todos os preditores, é preciso que as variâncias associadas aos efeitos aleatórios ou componentes de variância sejam conhecidas. Como na prática dispõe-se apenas de estimativas para essas variâncias, essas predições são denominadas de Empirical-BLUP (LITTELL, 2002).

No método estatístico geral de seleção sob modelo misto de Henderson, a predição do efeito de genótipo, no qual centra-se, em geral, o interesse do melhorista, é efetuada uma penalização (encolhimento) da performance média fenotípica do genótipo, ajustada para os demais efeitos ambientais controlados presentes no experimento. O primeiro trabalho conhecido que utilizou tal preditor foi o de Lush (1931). Nesse trabalho, o autor define uma técnica preditiva do valor reprodutivo do touro para produção de leite e gordura em função do desempenho médio de suas filhas. O preditor utiliza a correlação entre o genótipo do touro e a média de produção das filhas. De acordo com Lush (1931), a correlação mede a tendência de duas coisas variarem juntas e, portanto serem previsíveis uma em relação à outra. Essa correlação indica o quão fidedigno é o valor reprodutivo do touro predito pela performance média de sua progênie. Posteriormente, algumas técnicas foram desenvolvidas seguindo o mesmo princípio. Dentre elas pode-se citar o preditor de Bayes, o índice multiefeitos (LUSH, 1947) e o estimador de James-Stein (EFRON; MORRIS, 1977). Atualmente, o preditor de Henderson (HENDERSON, 1975) - BLUP - é o mais difundido (PIEPHO et al., 2008; RESENDE, 2007).

Para se proceder a modelagem segue-se dois passos. Primeiro modela-se a média dos efeitos fixos e posteriormente a covariância dos aleatórios. Modelar a média significa construir as matrizes de incidência dos fatores presentes no modelo de forma a se ter a mínima diferença entre o valor predito pelo modelo e o observado. Para os fatores aleatórios modela-se também a estrutura de covariância. Esse processo nada mais é que o modo como as variâncias e covariâncias irão se comportar ao se predizerem os efeitos da variável aleatória. Ambas as modelagens são de fundamental importância na experimentação e necessitam ter bom ajustamento e parcimônia (ARNAU et al., 2010; CNAAN; LAIRD; SLASOR, 1997; LITTELL; PENDERGAST; RANJINI, 2000; VALLEJO et al., 2011).

O ajustamento do modelo é verificado pela verossimilhança, a qual informa a capacidade do modelo em interpretar a realidade, de ser verossímil. Ou seja, indica a probabilidade do modelo escolhido ser mais adequado aos dados. Essa estimativa é função do número de parâmetros adicionados ao modelo. Seguindo-se esse raciocínio o incremento do número de parâmetros no modelo conduz ao aumento da verossimilhança. Quanto mais parâmetros, melhor o ajustamento. Porém a questão não é unilateral. O excesso de parâmetros a serem estimados compromete a capacidade preditiva do modelo. O problema agrava-se quanto menor a quantidade de dados, ou equações, disponíveis na estimação. Para contornar esse problema, alguns pesquisadores desenvolveram os chamados critérios de informação. Esses critérios penalizam a verossimilhança em função do número de parâmetros a serem estimados e, dependendo do critério, também em relação ao tamanho amostral (AKAIKE, 1974; BURHAN; ANDERSON, 2004; SCHWARZ, 1978).

Os critérios de informação podem ser divididos em consistentes e eficientes. Critérios eficientes escolhem o melhor modelo de dimensão finita quando o verdadeiro modelo, que é desconhecido, é de dimensão infinita. Já os critérios consistentes escolhem o modelo correto, com probabilidade se aproximando da unidade, quando o modelo verdadeiro de dimensão finita é considerado existente. A maioria dos critérios de informação encontra-se em uma dessas duas categorias. Os mais utilizados são os de Akaike (AIC), Akaike corrigido (AICC) e o critério de Schwarz (BIC). Consideram-se os dois primeiros como eficientes e o último como critério consistente (GURKA, 2006).

O AIC penaliza a verossimilhança de acordo com o número de parâmetros, idêntico ao que se faz no BIC. Porém, o último considera também o tamanho amostral em estudo. Um critério é consistente em ordem quando não varia sua resposta com o tamanho amostral. O AIC tem sido criticado em função disso, o que não ocorre com o BIC. Além desse fato, o AIC tende ao excesso de

ajustamento. Para contornar a falha, Burham e Anderson (2004) propuseram uma correção para o AIC. Esse critério de Akaike corrigido (AICC) supera o não corrigido em amostras pequenas. O critério de Schwarz derivou-se de uma solução de Bayes para o problema da identificação de modelos. O valor de BIC é minimizado assintoticamente na ordem do modelo que apresenta a maior probabilidade posterior. Esses critérios são interpretados verificando-se o valor calculado. Quanto menor o valor, menor o erro de predição e melhor o modelo (BURHAN; ANDERSON, 2004; KNAFL; BEEBER; SCWARTZ, 2012; KOEHLER; MURPHREE, 1988).

Os critérios de seleção de modelos citados acima, não garantem a escolha do modelo verdadeiro, apenas indicam qual modelo é o mais provável dentre os analisados e referentes aos dados em questão. As respostas dos diferentes critérios de informação e seleção de modelos variam em função de algumas características, como a distribuição de probabilidade dos dados e, principalmente do padrão de covariância presente nos mesmos. Outra nuance presente é o tamanho amostral. Alguns estudos já foram realizados com o intuito de se identificar o melhor critério de informação. Tais estudos podem ser realizados por simulação e por meio de dados reais. A diferença é que na simulação o verdadeiro modelo é assumido conhecido. E por isso, a comparação dos critérios de escolha é feita analisando-se a proporção de vezes que o método foi capaz de identificar o verdadeiro modelo. Contrariamente, a comparação com dados reais é realizada apenas verificando-se a proporção de coincidências dos critérios sob análise ao escolher o melhor modelo (LIU; ROVINE; MOLENAAR, 2012).

Para investigar o desempenho dos critérios de informação mais utilizados na seleção de modelos estatísticos, Vallejo et al. (2011) obtiveram resultados interessantes. Os autores usaram tamanhos de amostras reais e padrões de covariâncias complexos. Outra contribuição foi a de testar o

desempenho dos critérios de informação quando a suposição de normalidade é violada. Os autores simularam a combinação entre duas modelagens de médias com quatro estruturas de covariâncias. Variou-se também o número de sujeitos e de medições. Foram gerados dados simulados das 192 condições manipuladas. Para cada condição foram ajustados em média 36 modelos. Os dois principais resultados encontrados pelos autores dizem respeito à não consistência na escolha do modelo correto pelos critérios de informação testados e ao fato de que não há superioridade uniforme entre eles. No entanto, critérios consistentes, em geral, comportaram-se melhor do que os eficientes. Critérios consistentes, como o BIC, desempenharam-se melhor sob estruturas de covariâncias simples comparando-se com estruturas mais complexas. Esse padrão é semelhante tanto sob normalidade dos dados como em sua falta. Entretanto há uma clara superioridade na acurada detecção de modelos quando os dados provêm de uma distribuição normal.

2.2.2 Análise de dados longitudinais

Dados longitudinais são obtidos por repetidas mensurações de um caráter em um mesmo indivíduo ou na mesma parcela. O termo dados longitudinais é uma generalização de medidas repetidas. A característica determinante nesse tipo de estudo é o padrão de covariâncias existente nos fatores aleatórios do modelo estatístico. Esses padrões provêm da biologia do caráter em estudo. A consequência direta é a não independência entre as medições. Existe correlação entre elas. Há três razões principais para a realização de estudos com medidas repetidas (COLLINS, 2006). A primeira delas é a obtenção de estimativas precisas dos parâmetros populacionais. Outra razão é o monitoramento da variável resposta, embora o foco principal esteja nos

efeitos de tratamentos. Por último, esses estudos visam avaliar o efeito dos fatores no decorrer do tempo.

No melhoramento, o principal objetivo da análise longitudinal é separar da variação total contida nos dados a porção devida às diferenças entre indivíduos, que é genética e pode ter um componente de ambiente permanente. Dessa forma, a variância dentro de indivíduos, também chamada por alguns autores de plasticidade fenotípica, não confundirá a identificação de indivíduos superiores (ALBERT, 1999; COLLINS, 2006; DIGGLE et al., 2002; POL; VERHULST, 2006).

A experimentação com medidas repetidas é comum, especialmente na pesquisa agrícola com plantas perenes. Todavia, independente da área de pesquisa, as questões concernentes à análise de experimentos com medidas repetidas são sempre as mesmas (CHENG et al., 2010). Como modelar corretamente? Como tratar a estrutura de covariância dos efeitos aleatórios do modelo?

Basicamente, há nesse tipo de estudo uma gradação entre a Análise de Variância com medidas repetidas (ANAVA-R) e a Análise Multivariada. Há também a ANAVA tradicional que considera independência entre as medições, que não é indicada para esse tipo de análise. Essa estrutura é chamada de componentes de variância (VC). No estudo multivariado não existe qualquer padrão preestabelecido para as covariâncias. A matriz de covariância é dita não estruturada (UN), podendo seus elementos atingir quaisquer valores. Esse tem sido considerado o procedimento ótimo para a análise de medidas repetidas. No entanto, o excesso de parâmetros a serem estimados pode comprometer a análise reduzindo sua confiabilidade. Alternativamente, tem-se a ANAVAR, procedimento amplamente difundido no contexto e utilizado há tempos na pesquisa. Existem várias denominações para esse delineamento, sendo o mais comum o termo parcelas subdivididas no tempo e modelo completo de

repetibilidade (APIOLAZA; GARRICK, 2001; EVERITT, 1999; KNAFL; BEEBER; SCHWARTZ, 2012; RESENDE, 2007).

Na análise utilizando-se o modelo de parcelas subdivididas no tempo tem-se a estrutura de covariância denominada de simetria composta (CS). Esse padrão postula que as variâncias de cada medição são iguais, bem como a correlação existente entre elas. Ou seja, as medições são igualmente correlacionadas independentemente da distância temporal entre elas. Tais restrições são consideradas severas e somente são eficientes quando a esfericidade é atingida. Esfericidade diz respeito à igualdade da variância da diferença entre as médias dos níveis de fatores estudados. Essa condição é facilmente verificada por meio do teste de esfericidade de Mauchly. Caso a hipótese nula, que atesta a esfericidade for rejeitada, os testes de F não podem ser diretamente aplicados. Nesse caso, há incremento na taxa de erro tipo I. Entretanto, existem algumas opções na literatura. Dentre elas cita-se a correção dos graus de liberdade de Greenhouse e Geisser (1959) e Huynh e Feldt (1976). Porém esse tipo de ajustamento não tem se mostrado satisfatório em função dos diferentes padrões de covariâncias existentes (EYDURAN; AKBAS, 2010; GLESER, 1966; HUYNH; FELDT, 1970, 1976; KESELMAN; ALGINA; KOWALCHUK, 2001).

Uma abordagem flexível consiste em modelar a estrutura de covariâncias dos dados de modo a promover maior precisão (CHENG et al., 2010). As opções variam da completa independência entre as medições, em que os valores de covariância são zeros, até um padrão não estruturado. Em UN os valores de covariância podem assumir qualquer valor. Estruturas intermediárias, como por exemplo, a autorregressiva de primeira ordem (AR) considera que as correlações são decrescentes à medida que o intervalo de tempo aumenta entre elas. Algumas estruturas possuem a vantagem adicional de considerarem a heterogeneidade de variâncias. Pode-se, dessa forma, lançar mão de estruturas

de covariâncias mais condizentes com o comportamento do caráter em estudo, obtendo-se uma boa qualidade de ajuste, aliado à manutenção da parcimônia do modelo. Os modelos para análise de medidas repetidas no melhoramento genético apresentam, além do erro experimental, outros fatores de efeitos aleatórios, a exemplo da interação dos genótipos com os ambientes em teste. Portanto, modelam-se, nessa situação, matrizes de covariâncias relacionadas aos erros (R) e ao efeito de genótipos em cada medição ou colheita, por meio da matriz da interação progênes com as colheitas (PC).

No melhoramento de plantas, a modelagem da estrutura de covariâncias dos fatores aleatórios não é comum. Em plantas perenes, a abordagem vem ganhando espaço em relação aos métodos tradicionais em várias culturas. Em *Coffea canephora* Cilas, Montagnon e Bar-Hen (2011) estudaram a estabilidade produtiva de vinte clones na estação experimental de Divo, no Centro Nacional de Pesquisas Agronômicas (CNRA). O delineamento utilizado foi o de parcela de árvore única aleatorizado com dois blocos. Foram usados trinta réplicas por clone, sendo quinze por clone e quinze por bloco. Clones foram considerados de efeito aleatório e blocos como fixos. Analisaram-se dois ciclos produtivos. O primeiro foi composto das quatro primeiras colheitas anuais antes da poda e o segundo ciclo continha as cinco colheitas após a poda. Apesar do objetivo principal da pesquisa ser a estabilidade produtiva, os autores avaliaram o ajustamento de cinco estruturas de covariâncias, a saber; CS, CSH, AR, ARH e UN, Para as nove colheitas os autores identificaram a CSH como a estrutura adequada para os dados. Em função da estrutura assumida pela covariância realizaram interessantes interpretações em relação à influência ambiental no caráter.

Em outro estudo com *Coffea canephora*, Cecon et al. (2008) analisaram dados de produção de cinquenta clones de conilon durante cinco colheitas, empregando diferentes matrizes de covariâncias dos erros. Por meio de AIC

identificaram que a estrutura de simetria composta com variâncias heterogêneas (CSH) foi a que proporcionou o melhor ajustamento. No entanto, trabalhos envolvendo a modelagem de matrizes de covariância além da residual são escassos. Esse fato não se observa somente no melhoramento, como também em outras áreas da pesquisa (LIU; ROVINE; MOLENAAR, 2012).

Para propor uma abordagem para a análise de produtividade em clones de cana-de-açúcar, melhoristas australianos descrevem uma modelagem que combina informações referentes às colheitas em um mesmo experimento quanto em diferentes ambientes (SMITH et al., 2007). A questão-chave são as diferentes modelagens para a covariância genética (G) e residual (R). Foram utilizadas duas bases de dados para exemplificação. A primeira refere-se a seis experimentos de campo nos quais foram avaliados um total de 240 clones. A segunda é composta de cinco experimentos em diferentes localidades. No total, 112 clones foram plantados. A matriz R foi modelada combinando-se informações temporais e espaciais. A obtenção de R é por meio dos produtos de Kronecker das matrizes que contêm as correlações temporais com as que acomodam correlações de colheitas e experimentos. Similarmente à modelagem da G, considerou-se a covariância de colheitas e experimentos, as quais assumiram diferentes estruturas. Ajustou-se primeiro a R e posteriormente a G. O melhor ajustamento encontrado para a primeira base de dados por meio de AIC foi, para a matriz R, uma combinação entre UN e AR. Enquanto para a G foi a UN e Fator Analítica de Segunda Ordem (FA). Para o exemplo dois, a modelagem residual escolhida foi a mesma do primeiro. Entretanto para a matriz G o padrão UN não convergiu. Ficando apenas a estrutura fator analítica com dois termos multiplicativos FA como o melhor ajustamento. Uma conclusão importante dos autores foi referente à predição do valor genético dos clones pelo método tradicional quando comparado ao utilizado em seu trabalho. Os valores encontrados por seu método conduzem a diferentes seleções, além de

promoverem estimativas mais acuradas. A diferença entre o quadrado médio do erro de predição entre elas foi de 19% para o primeiro exemplo e 39% no segundo.

Outro estudo foi realizado em Pinus (*Pinus radiata*) em que 45 progênies de meios irmãos (MI) foram avaliadas juntamente com nove testemunhas. Objetivou-se testar o ajustamento das modelagens de covariância e verificar seu efeito na estimação de parâmetros populacionais. O caráter em estudo foi altura de planta, a qual foi mensurada em 10 idades diferentes, de um a quinze anos. A matriz R foi considerada UN enquanto foram testadas várias modelagens para G. Para os blocos considerou-se uma estrutura AR de primeira ordem, bem como a matriz de covariância dos efeitos de parcela. Os autores encontraram resultados interessantes. A melhor modelagem de G foi a AR. Essa estrutura promoveu alterações nos valores de herdabilidade e nas correlações genéticas entre as medições. Os autores concluem o trabalho sugerindo que a estrutura de covariância é específica do caráter. Varia também em função do número de progênies avaliadas e dos erros de amostragem. Dessa forma, diferentes caracteres requerem estruturas específicas, embora a estrutura autorregressiva pareça ser flexível o bastante para ser utilizada no melhoramento de espécies afins (APIOLAZA; GILMOUR; GARRICK, 2000).

Recentemente, pesquisadores brasileiros procuraram a melhor forma de se modelar o caráter número de frutos por planta. Objetivou-se comparar formas de análise de medidas repetidas para o melhoramento genético da Pinha. Utilizaram-se diferentes estruturas de covariância para os fatores progênie, parcela e resíduo. Avaliaram-se vinte progênies de MI durante três colheitas anuais. Compararam-se as modelagens de repetibilidade (RE); repetibilidade mais interação genótipos por ambientes (REGA); ARH; ANTE e UN. A comparação dos modelos foi realizada por meio de teste de qui-quadrado (χ^2) e AIC. As melhores estruturas de covariâncias encontradas pelos autores foram a

ANTE para o fator progênie e de parcela e UN para o resíduo. Concluíram ainda que a escolha do melhor modelo é necessário para a análise de medidas repetidas na Pinha para se obter eficiência no melhoramento (MARIGUELE et al., 2011).

3 MATERIAL E MÉTODOS

O experimento foi conduzido pela Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG) na Fazenda Experimental de Machado, MG (latitude de 21° 40'S, a longitude de 45° 55'W). A área experimental apresenta solo classificado como Latossolo Vermelho Distróférrico, com relevo ondulado. A altitude é de 881 m, a precipitação anual de 1670mm e a temperatura média de 21°C.

Foram avaliadas 21 progênies endogâmicas $S_{0,1}$ provenientes dos cruzamentos de Mundo Novo com Mundo Novo e Mundo Novo com Bourbon Amarelo (Tabela 1), desenvolvidas pelo programa de melhoramento genético do cafeeiro do Insitituto Agrônômico de Campinas (IAC), SP. O experimento foi instalado em 15 de janeiro de 1988, no qual foram avaliadas 8 colheitas. O delineamento experimental utilizado foi o de blocos casualizados, com três repetições. Cada parcela foi constituída de uma linha de oito plantas, no espaçamento de 3,0 x 1,5 m.

As calagens e adubações de solo e foliares foram realizadas ao longo de toda a condução do experimento, conforme a necessidade da cultura. Os tratos fitossanitários foram realizados preventivamente ou curativamente, de acordo com a sazonalidade da ocorrência das pragas e doenças.

Tabela 1 Relação de progênies $S_{0.1}$ provenientes dos cruzamentos entre linhagens de Mundo Novo com Mundo Novo e Mundo Novo com Bourbon Amarelo avaliadas em Machado, Minas Gerais

Codificação	Progênies
1	474 – 323 – C 562 – LCH 6682
2	475 – 324 – C 640 – LCH 6683
3	476 – 322 – C 472 – LCH 6680
4	477 – 320 – C 715 – LCH 6677
5	478 – 315 – C 790 – LCH 6667
6	479 – 316 – C 251 – LCH 6669
7	480 – 309 – C 452 – LCH 1596
8	481 – 326 – C 383 – LCH 6698
9	482 – 319 – C 163 – LCH 6675
10	483 – 308 – C 269 – LCH 1595
11	484 – 313 – C 122 – LCH 6653
12	485 – 321 – C 619 – LCH 6679
13	486 – 325 – C 76 – LCH 6684
14	487 – 318 – C 21 – LCH 6674
15	488 – 317 – C 190 – LCH 6672
16	489 – 312 – C 714 – LCH 3223
17	490 – 311 – C 757 – LCH 2980
18	491 – 310 – C 178 – LCH 2917-3
19	492 – 314 – C 584 – LCH 6664

20	493 – 327 – C 754 – LCH 6705
21	490– 62 – C 669 – LH 2931 P17 III

3.1 Análise dos dados

Realizou-se a análise exploratória dos dados para verificar as pressuposições do modelo estatístico (RAMALHO; FERREIRA; OLIVEIRA, 2012).

Primeiramente foram realizadas as análises individuais por colheita, por meio do procedimento PROC MIXED do programa SAS ® (STATISTICAL ANALYSIS SYSTEM INSTITUTE - SAS INSTITUTE, 2009). O modelo estatístico utilizado foi o modelo a seguir (1).

$$y_{ij} = m + b_j + p_i + e_{ij} \quad (1)$$

em que:

y_{ij} : observação da ij-ésima parcela no bloco j que recebeu a progênie i;

m : constante associada a todas as observações;

b_j : efeito fixo do o j-ésimo bloco;

p_i : efeito aleatório da i-ésima progênie, sendo $p_i \sim \text{NMV}(0, \sigma_p^2)$;

e_{ij} : efeito aleatório do erro experimental associado à observação da ij-ésima parcela, sendo $e_{ij} \sim \text{NMV}(0, \sigma_e^2)$.

A partir das análises individuais foram obtidas as estimativas dos coeficientes de variação ambiental (CV) e das herdabilidades de acordo com Ramalho et al. (2012). Além disso, foram computadas as predições E-BLUP

(Empirical Best Linear Unbiased Prediction) dos valores genotípicos das progênies (LITTELL et al., 2006).

Para construção dos gráficos Box Plot das previsões E-BLUP das progênies, por colheita e em biênio, utilizou-se o programa R (R DEVELOPMENT CORE TEAM, 2008). As correlações classificatórias de Spearman dos E-BLUP das progênies foram calculadas no programa SAS, por meio do proc CORR.

Posteriormente foi realizada a análise no esquema de parcelas subdivididas no tempo para cada modo de se considerar os dados (STEEL; TORRIE; DICKEY, 1997). Os agrupamentos dos dados foram realizados analisando-se as 8 colheitas anuais, em biênios, anos bons (colheitas 2, 4, 6 e 8) e anos ruins (colheitas 1, 3, 5 e 7). Utilizou-se o seguinte modelo estatístico.

$$y_{ijq} = m + b_j + p_i + c_q + pb_{ij} + bc_{jq} + pc_{iq} + e_{ijq} \quad (2)$$

y_{ijq} : observação da ijq -ésima parcela no bloco j na colheita q que recebeu a progênie i ;

m : constante associada a todas observações;

b_j : efeito fixo do j -ésimo bloco;

p : efeito aleatório da i -ésima progênie, sendo $p_i \sim \text{NMV}(0, \sigma_p^2)$;

c_q : efeito fixo da q -ésima colheita;

pb_{ij} : efeito aleatório da ij -ésima interação progênie com bloco, sendo

$pb_{ij} \sim \text{NMV}(0, \sigma_{pb}^2)$;

bc_{jq} : efeito fixo da jq -ésima interação bloco com colheita;

pc_{iq} : efeito aleatório da iq -ésima interação progênie com colheita,

sendo $pc_{iq} \sim \text{NMV}(0, \sigma_{pc}^2)$;

e_{ijq} : efeito aleatório do erro experimental associado à observação da ijq -ésima parcela, sendo $e_{ijq} \sim \text{NMV}(0, \sigma_e^2)$.

O modelo de parcela subdividida no tempo na forma matricial é denotado por:

$$y = X_m + Z_{Bi}b + Z_{Pi}pc + Z_{Ti}p_i + e_i \quad (3)$$

X_m : matriz de incidência e vetor m de feitos fixos;

$Z_{Bi}b$: matriz de incidência e vetor b de efeitos aleatórios da interação bloco com progênie;

$Z_{Pi}pc$: matriz de incidência e vetor pc de efeitos aleatórios da interação progênie com colheita;

$Z_{Ti}p_i$: matriz de incidência e vetor de p_i de efeitos aleatórios de progênie;

e_i : vetor a de erros aleatórios.

Os valores esperados e a matriz de dispersão assumindo distribuição normal multivariada (NMV) são:

$$\begin{bmatrix} y \\ b \\ pc \\ p_i \\ e_i \end{bmatrix} \sim \text{NMV} \begin{bmatrix} X_m \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} V_i & Z_{Bi}B_0 & Z_{Pi}PC_0 & Z_{Ti}P_0 & R_0 \\ B_0Z_{Bi} & B_0 & 0 & 0 & 0 \\ PC_0Z_{Pi} & 0 & P_0 & 0 & 0 \\ PZ_{Ti} & 0 & 0 & G_0 & 0 \\ R_0 & 0 & 0 & 0 & R_0 \end{bmatrix}$$

B_0 , PC_0 , P_0 , R_0 são as matrizes de covariância para os fatores interação bloco com progênie, interação progênie com colheita, progênie e resíduo respectivamente. 0 representa matriz nula.

Verificou-se a hipótese da esfericidade da matriz de covariâncias residuais por meio do teste da esfericidade de Mauckhly (MAUCKHLY, 1940) pelo proc GLM do SAS® (SAS INSTITUTE, 2009). Após a rejeição da esfericidade foram testadas onze estruturas de matrizes de covariâncias, desde a mais simples, simetria composta (CS) até a não estruturada (UN), considerada a mais complexa. As análises das estruturas de covariâncias foram realizadas pelo proc MIXED dos SAS®.

A escolha da estrutura mais adequada para (R) e (PC) foi obtida pela ponderação do valor obtido pelo critério de informação de Schwarz (BIC). A estimativa de BIC é obtida por $-2 \log L + p \log v$, em que L é o valor estimado da verossimilhança do modelo, p é o número de parâmetros estimados e v é o número de graus de liberdade do resíduo. O cálculo da ponderação foi realizado conforme proposto por Gonçalves e Fritzche-Neto (2012). O resultado obtido (w_i) informa qual a probabilidade do modelo escolhido ser o correto em relação aos considerados nas comparações. A expressão utilizada para a ponderação foi:

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{i=1}^r \exp(-\Delta_i/2)} * 100 \quad (5)$$

w_i : probabilidade do modelo i ser o correto em relação aos demais utilizados na comparação. Δ_i é a diferença entre o menor valor de BIC (melhor ajuste) dentre os analisados com o valor obtido por cada modelo considerado. O menor valor foi utilizado, pois os critérios de informação calculados pelo SAS®

o consideraram como o melhor. Ou seja, menor valor de BIC maior ajustamento do modelo aos dados. De acordo com Gonçalves e Fritzsche-Neto (2012) esse cálculo é apenas uma mudança de escala para comparar as diferenças de BIC entre os modelos em uma escala de 0-1.

A escolha da melhor estrutura para as matrizes de covariância foi sequencial. Primeiramente identificou-se a estrutura mais provável para (R) e posteriormente para (PC). Para isso, testaram-se todas as estruturas para R e com os valores calculados por meio da expressão (5) escolheu-se a mais provável. Com a estrutura identificada como a mais adequada para (R) repetiu-se o processo para a matriz de covariância (PC). Estruturas de matrizes de covariâncias que apresentaram problemas de convergência não foram consideradas nas análises.

3.2 Estruturas de covariância

As estruturas das matrizes de covariância (C_{ov}) apresentadas serão exemplificadas considerando-se quatro colheitas apenas. Na diagonal principal das matrizes encontram-se os parâmetros relacionados às variâncias de cada avaliação. Fora da diagonal os de covariância para cada par de colheitas considerado. O número de parâmetros estimados por cada estrutura de covariância é função da dimensão global (t) da matriz de covariância. Para a modelagem da matriz R, a variável (t) corresponde ao número de colheitas e para PC ao número de interações entre as progênies e as colheitas. Os parâmetros estimados nas matrizes são: a variância (σ^2), o desvio padrão (σ) e a correlação (ρ).

3.2.1 Simetria composta (CS)

A estrutura pressupõe a inexistência de heterogeneidade de variâncias e que a correlação para cada par de colheitas é constante. Essa estrutura utilizada na matriz R equivale ao esquema de parcelas subdivididas no tempo. São necessários estimar dois parâmetros, a variância (σ^2) e desvio padrão (σ) do fator em estudo.

$$C_{ov} = \begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

3.2.2 Antedependência (ANTE)

Na estrutura de Antedependência de primeira ordem assume-se que uma observação no tempo considerado pode ser explicada por observações prévias. Nesta estrutura são necessárias a estimação de $2t-1$ parâmetros.

$$C_{ov} = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_1 & \sigma_1 \sigma_3 \rho_1 \rho_2 & \sigma_1 \sigma_4 \rho_2 \rho_3 \\ \sigma_2 \sigma_1 \rho_1 & \sigma_2^2 & \sigma_2 \sigma_3 \rho_2 & \sigma_2 \sigma_4 \rho_3 \rho_4 \\ \sigma_3 \sigma_1 \rho_2 \rho_1 & \sigma_3 \sigma_2 \rho_2 & \sigma_3^2 & \sigma_3 \sigma_4 \rho_3 \\ \sigma_4 \sigma_1 \rho_3 \rho_2 & \sigma_4 \sigma_2 \rho_4 \rho_3 & \sigma_4 \sigma_3 \rho_3 & \sigma_4^2 \end{bmatrix}$$

Em que:

σ_i^2 : refere-se ao i-ésimo parâmetro de variância;

ρ_k : é o i-ésimo parâmetro de autocorrelação, satisfazendo a condição

de $|\rho_k| < 1$.

3.2.3 Autorregressiva (AR)

Na estrutura autorregressiva de primeira ordem considera-se homogeneidade de variâncias e que as correlações entre colheitas mais distantes no tempo são menores. São estimados dois parâmetros, a correlação autorregressiva (ρ) e a variância.

$$C_{ov} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

3.2.4 Autorregressiva com variâncias heterogêneas (ARH)

Específica a estrutura autorregressiva com variâncias heterogêneas. Resende (2007) comenta que, para intervalos iguais entre as medições, esse modelo se assemelha ao processo caráter de Pletcher e Geyer (1999). Eles foram desenvolvidos para reduzir a imposição de estacionariedade das correlações. São necessários se estimar t-1 parâmetros.

$$C_{ov} = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho & \sigma_1 \sigma_3 \rho^2 & \sigma_1 \sigma_4 \rho^3 \\ \sigma_2 \sigma_3 \rho & \sigma_2^2 & \sigma_2 \sigma_3 \rho & \sigma_2 \sigma_4 \rho^2 \\ \sigma_3 \sigma_1 \rho^2 & \sigma_3 \sigma_2 \rho & \sigma_3^2 & \sigma_3 \sigma_4 \rho \\ \sigma_4 \sigma_1 \rho^3 & \sigma_4 \sigma_2 \rho^2 & \sigma_4 \sigma_3 \rho & \sigma_4^2 \end{bmatrix}$$

3.2.5 Médias móveis autorregressiva (ARMA)

Congrega características da estrutura autorregressiva de primeira ordem com a técnica das médias móveis (RESENDE; STURION, 2003). Nesse modelo estimam-se três parâmetros.

$$C_{ov} = \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}$$

Em que:

(ρ) é o parâmetro autorregressivo;

(γ) modela o componente de médias móveis.

3.2.6 Simetria composta com variâncias heterogêneas (CSH)

Estrutura de simetria composta com variâncias heterogêneas. Possui parâmetro diferente de variância para cada elemento da diagonal principal e utiliza a raiz quadrada desses parâmetros para compor as caselas fora da diagonal. Estima-se t+1 parâmetros.

$$C_{ov} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

3.2.7 Fator Analítica (FA)

Estrutura fator analítica com um fator (q) (FA 1). Esse modelo utiliza a técnica das variáveis latentes não observáveis (fatores). Considera que uma observação é uma combinação de fatores, ponderados pela intensidade com que o fator explica a variável observável (carga do fator), mais um erro de predição. Trata-se de uma técnica reducional da dimensão dos dados. Seu princípio é semelhante ao modelo de fatores principais aditivos e de interação multiplicativos (AMMI) utilizado para modelar a interação genótipos por ambientes (RESENDE, 2007). A diferença básica entre eles é que, em AMMI, modelam-se as variâncias e em FA as covariâncias. Além disso, FA é invariante às mudanças de escala nos eixos dos vetores, mas não à rotação dos mesmos como em AMMI. Estima-se $[q/2](2t-q+1)+t$ parâmetros (q refere-se ao número de vetores (eixos) utilizados para captar a variabilidade dos dados).

$$C_{ov} = \begin{bmatrix} \lambda_1^2 + d_1 & \lambda_1 \lambda_2 & \lambda_1 \lambda_3 & \lambda_1 \lambda_4 \\ \lambda_2 \lambda_1 & \lambda_2^2 + d_2 & \lambda_2 \lambda_3 & \lambda_2 \lambda_4 \\ \lambda_3 \lambda_1 & \lambda_3 \lambda_2 & \lambda_3^2 + d_3 & \lambda_3 \lambda_4 \\ \lambda_4 \lambda_1 & \lambda_4 \lambda_2 & \lambda_4 \lambda_3 & \lambda_4^2 + d_4 \end{bmatrix}$$

λ : desvio padrão da carga do fator. Essa estimativa se altera entre as colheitas, mas não entre progênies na mesma medição;

d : variância do erro entre a variável observável e o fator.

3.2.8 Huynh-Feldt (HF)

Esta matriz apresenta variâncias diferentes para cada colheita e covariâncias obtidas pela média aritmética das covariâncias (HUYNH; FELDT, 1970). Estima-se t+1 parâmetros.

$$C_{ov} = \begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_2^2 + \sigma_1^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_3^2 + \sigma_1^2}{2} - \lambda & \frac{\sigma_3^2 + \sigma_2^2}{2} - \lambda & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_4^2 + \sigma_1^2}{2} - \lambda & \frac{\sigma_4^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_4^2 + \sigma_3^2}{2} - \lambda & \sigma_4^2 \end{bmatrix}$$

λ : autovalor da matriz de covariância.

3.2.9 Toeplitz (TOEP)

Modelo de correlação bandada Toeplitz. Pode ser visto como um modelo de series temporais. Possui variâncias iguais para cada colheita e covariâncias diferentes em ocasiões não adjacentes. Estimam-se t parâmetros.

$$C_{ov} = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

3.2.10 Toeplitz com Variâncias Heterogêneas (TOEPH)

Representa a estrutura Toeplitz com variâncias heterogêneas. Estima-se $2t-1$ parâmetros.

$$C_{ov} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix}$$

3.2.11 Não estruturada (UN)

Permite que os parâmetros de variância e covariância assumam quaisquer valores. É a mais parametrizada, sendo necessárias a estimação de $t(t+1)/2$ parâmetros.

$$C_{ov} = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

Testou-se também o ajustamento do modelo e as inferências realizadas pelo modelo de parcela subdividida no tempo com transformação para heterogeneidades de variância, como proposto por Resende (2007). Essa correção é feita por um fator que multiplica os dados de cada parcela. Esse fator

foi obtido pela relação entre a raiz da herdabilidade ao nível de parcela da colheita em consideração pela raiz das médias das herdabilidades de todas as colheitas (h_i/h_m), e essa análise foi denominada de CS*.

3.3 Estimação de parâmetros

Para realizar a comparação da estimação de parâmetros, alguns modelos foram escolhidos em função de dois aspectos. O primeiro aspecto incluiu as comparações dos modelos considerados como os mais adequados aos dados pela expressão (5) para (R) e também aqueles no qual se modelou (R) e (PC). O segundo aspecto utilizou as comparações dos modelos que representam as análises em parcelas subdivididas no tempo com e sem transformação para heterogeneidade de variâncias (h_i/h_m); isso devido ao fato de serem as mais simples e comumente utilizadas.

Foram estimados 4 parâmetros. A herdabilidade na média das progênes (\hat{h}^2) conforme Ramalho et al. (2012); a acurácia na média das progênes (\hat{h}) como demonstrado por Resende (2007) e o ganho com a seleção (GS), que foi calculado com base na média dos E-BLUP's das cinco melhores progênes selecionadas.

Herdabilidade na média das progênes para simetria composta

$$\hat{h}^2 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \frac{\hat{\sigma}_{pc}^2}{q} + \frac{\hat{\sigma}_{pb}^2}{j} + \frac{\hat{\sigma}_e^2}{qj}}$$

Herdabilidade na média de progênes para as modelagens alternativas à CS

$$\hat{h}^2 = \frac{E - BLUP}{(\bar{Y}_{i..} - \bar{Y}_{...})} \quad (\text{RESENDE, 2007})$$

sendo $\bar{Y}_{i..}$ a média da progênie i em todas as colheitas e $\bar{Y}_{...}$ a média geral de todas as colheitas em consideração. Dessa forma, tem-se uma estimativa de herdabilidade para cada progênie das quais se extraiu a média aritmética.

Acurácia na média das progênes

$$\hat{h} = \sqrt{\hat{h}^2}$$

$\hat{\sigma}_p^2$: variância de progênie;

$\hat{\sigma}_{pc}^2$: variância da interação progênie com colheita;

$\hat{\sigma}_{pb}^2$: variância da interação progênie com repetição;

$\hat{\sigma}_e^2$: variância do erro;

j : número de repetições;

q : número de colheitas ou biênios.

3.4 Correlação de Spearman e índice de coincidência

As correlações de Spearman e os índices de coincidência foram utilizados para demonstrar o impacto que as diferentes modelagens das

estruturas de covariância causaram na seleção dessas progênies por ocasião das análises de colheitas anuais ou em biênios antes e após a poda. Conforme mencionado anteriormente, o ordenamento das progênies foi obtido pelas estimativas E-BLUP para cada modelagem proposta. A obtenção da correlação de Spearman foi realizada pelo programa SAS ®. O índice de coincidência no ordenamento das progênies foi obtido de acordo com a expressão de Hamblin e Zimmerman (1986):

$$IC = \left(\frac{A - C}{B - C} \right)$$

A: número de progênies coincidentes entre as cinco selecionadas na comparação dois a dois;

B: número de progênies selecionadas.

C: quantifica o número de progênies coincidentes devido ao acaso. Utilizou-se para o cálculo 5%. Portanto, o número de progênies coincidentes devido ao acaso, neste trabalho foi de $0,05*23$, o que resulta em 1,15.

4 RESULTADOS

Os resultados iniciam-se por uma análise exploratória para demonstrar as características intrínsecas dos dados de produção de grãos de cafeeiro analisados. Posteriormente, demonstram-se como as colheitas (medições) realizadas repetidamente em uma mesma parcela experimental afetam a covariância dos fatores aleatórios do modelo estatístico e, por último, encontram-se os resultados que demonstram como as estruturas assumidas pelas matrizes de covariância afetam o melhoramento genético.

4.1 Análise exploratória

Ao se analisarem as colheitas anuais percebe-se que apenas nas colheitas 6 e 8 obteve-se significância no componente de variância de progênies a 5% de probabilidade. Entretanto, considerando-se os biênios apenas não se atestou a existência de variação genética significativa ($P < 0,05$) no primeiro biênio. Sendo assim, a herdabilidade não pode ser considerada existente nos casos em que esse componente foi não significativo (Tabela 2).

Percebeu-se heterogeneidade de variâncias ambiental e genética por meio dos valores estimados para o coeficiente de variação ambiental (CV) e herdabilidade, no nível de média de progênies (h^2), sendo essa mais acentuada em termos de colheitas individuais (Tabela 2). O CV variou entre as colheitas individuais de 15,48% (colheita 8) a 121,47% (colheita 7), enquanto que, em termos de biênios, as estimativas variaram de 13,59% (biênio 7-8) a 27,31% (biênio 1-2). Para as h^2 , as estimativas flutuaram de 18% (colheita 7) a 98% (colheita 8), enquanto que para biênios esses valores foram de 32% (biênio 1-2) e 87% (biênio 7-8). Por meio da correlação de Spearrman constatou-se que a

média fenotípica foi negativamente correlacionada com o CV (-0,73) e positivamente correlacionada com a herdabilidade (0,53). Isso quer dizer que houve tendência de médias baixas estarem associadas com altos CV's e baixas h^2 . Como exemplo, a colheita 7 foi a de menor média, na qual o CV foi o mais elevado e, por conseguinte, menor magnitude da h^2 (Tabela 2).

A alternância entre alta e baixa produção iniciou-se a partir da segunda colheita. Esse fenômeno é conhecido como bienalidade e ocorre com frequência na produção de grãos de café e dificulta a análise conjunta desse tipo de dados. No valor médio de CV e herdabilidade, os biênios se assemelharam aos anos de alta produção, os quais foram melhores quanto a essas estimativas. Observou-se que as colheitas de safra baixa (Anos Ruins) obtiveram o menor valor médio da herdabilidade. Observou-se também, nos anos pouco produtivos, média do CV superior aos demais casos (Tabela 2).

Tabela 2 Estimativa da média, do coeficiente de variação ambiental e da herdabilidade na média das progênies para colheitas individuais e em biênios

Colheita	Média		
	(Kg parcela ⁻¹)	CV (%)	h^2 (p-valor) ^{1/}
1	15,63	47,32	0,34 (0,1641)
2	64,77	25,37	0,40 (0,1168)
3	9,85	81,69	0,44 (0,0999)
4	53,24	28,80	0,43 (0,1028)
5	4,38	120,78	0,51(0,0650)
6	85,87	21,48	0,70 (0,0153)
7	3,56	121,47	0,18 (0,3000)
8	96,24	15,48	0,98 (0,0037)
1-2	40,20	27,31	0,32 (0,1737)

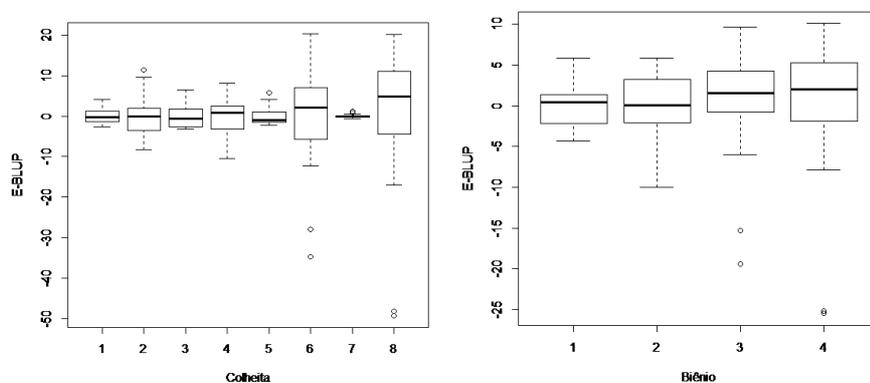


Figura 1 Box Plot dos E-BLUP's das colheitas anuais e agrupadas em biênios

Nota: O comprimento do retângulo representa o intervalo interquartil. O traço horizontal dentro do retângulo representa a mediana. A linha horizontal acima do retângulo demonstra o valor máximo e abaixo o valor mínimo. Os pontos acima do valor máximo e /ou abaixo do valor mínimo representam os outliers, os quais estão acima ou abaixo 1,5 vezes do quartil superior ou inferior.

Houve alteração no ordenamento dos E-BLUP's das progênies ao longo das colheitas avaliadas (Tabela 3). Isso especialmente foi evidenciado ao se correlacionarem colheitas de anos ruins com anos bons. As correlações foram mais altas e positivas entre colheitas de anos bons ou ruins, contudo maiores valores de correlações foram observados entre os anos mais produtivos (2, 4, 6 e 8). Notou-se, ainda, um aumento na magnitude da correlação ao se tomar colheitas mais avançadas como referência. A correlação entre a colheita 2 e a última colheita (8) foi de 0,62, enquanto que a correlação entre as colheitas 7 e 8 foi de 0,87.

Tabela 3 Correlações de Spearman dos E-BLUP's de progênies entre as colheitas anuais

Colheitas	2	3	4	5	6	7	8
1	0,48**	0,40**	0,41	0,19	0,42**	-0,04	0,49**
2		-0,03	0,65**	-0,16	0,67**	-0,15	0,62**
3			-0,03	0,62**	0,07	0,23	0,17
4				-0,32	0,88**	-0,53**	0,92**
5					-0,18	0,62**	-0,17
6						-0,37	0,87**
7							-0,51**

**Correlação significativa a 5%.

Em biênios observou-se que ao considerar ocasiões mais próximas no tempo incrementa-se a correlação entre eles (Tabela 4). O biênio 1 foi igualmente correlacionado com os demais, com uma estimativa de aproximadamente 0,68. Quando o biênio 2 é tido como referência a correlação eleva-se para 0,86 com o 3 e para 0,84 com o último biênio.

Tabela 4 Correlações de Spearman dos E-BLUP's de progênies entre os biênios

Colheitas	2	3	4
1	0,66**	0,68**	0,69**
2		0,86**	0,84**
3			0,86**

**Correlação significativa a 5%.

A observância das estimativas de herdabilidade ao longo das medições, sejam elas colheitas anuais ou em biênios, e das correlações auxiliam no

entendimento do comportamento das progênies e na definição de modelos estatísticos mais apropriados no que diz respeito à matriz de covariâncias dos resíduos (matriz R). Em geral, dados de colheitas são analisados conjuntamente admitindo a matriz R do tipo simetria composta (CS) ou modelo de parcela subdividida no tempo. Contudo, esse tipo de suposição pode não estar correta. Para a verificação da validade dessa suposição deve-se empregar previamente o teste de Mauchly.

A hipótese da esfericidade foi rejeitada pelo teste de Mauchly para colheitas anuais, biênios e anos ruins, indicando que a estrutura CS, para as matrizes R, pode não ser a mais adequada aos dados em questão. Dessa forma, estruturas alternativas da matriz R foram buscadas no intuito de lidar com a realidade presente, explorar melhor os dados e assim obter estimativas precisas. Quando se considerou a análise conjunta das colheitas de safra alta (Anos Bons) a esfericidade foi aceita, demonstrando que a simetria composta pode ser empregada com segurança (Tabela 5).

Tabela 5 Valores P do teste de esfericidade de Mauchly para as colheitas individuais, biênios, anos de alta (bons) e baixa (ruins) produtividade

Estratégias	Todas	Biênios	Anos Bons	Anos Ruins
p-valor	0,0001	0,0282	0,0911	0,0001

Dentre os modos de se analisar os dados, as colheitas anuais, biênios e anos bons apresentaram componentes de variância significativos a 5% para os fatores progênie e interação progênies com colheitas. Para os anos ruins, o fator genético foi significativo a 11% e o de interação progênies com colheitas a 6%, considerando-se a análise em esquema de parcela subdividida no tempo (Tabela 22, ANEXO A).

4.2 Estruturas de covariância

Partindo-se da rejeição da esfericidade, testou-se estruturas alternativas para as matrizes de covariância associada ao erro (R) e para a interação progênies com as colheitas (PC). Apesar de a análise conjunta dos anos bons terem apresentado esfericidade modelou-se também nessa condição para se analisar as consequências. A escolha da matriz foi baseada no critério de Schwarz (BIC).

Percebeu-se que o BIC indicou TOEP (correlação bandada) a estrutura para a matriz R mais adequada para as colheitas anuais, que considera correlações específicas para cada intervalo entre as colheitas. Para biênios, a estrutura escolhida foi a simetria composta com variâncias heterogêneas (CSH), a qual assume iguais correlações entre as medições lidando com a heterogeneidade de variâncias entre as colheitas consideradas. Para os anos de safra alta (anos bons) a estrutura autorregressiva com variâncias heterogêneas (ARH) foi escolhida, indicando correlações diferentes entre as colheitas consideradas e reconhecendo a existência de correlação serial, na qual medições mais próximas no tempo são mais correlacionadas. Em anos ruins o modelo que melhor se ajustou aos dados foi o antedependência estruturado (ANTE), o qual semelhantemente à ARH, assume que as correlações mais distantes são menos correlacionadas e ainda lidam com a heterogeneidade de variâncias (Tabela 6).

Tabela 6 Probabilidade (\hat{w}_i), em porcentagem, de o modelo da estrutura de covariância residual (R) ser o correto nos diferentes modos de se considerar os dados (anual, biênio, anos bons e anos ruins)

MODELO	ANNUAL	BIÊNIO	ANOS BONS	ANOS RUINS
CS	0,00	0,00	5,49	0,00
CS*	0,00	0,00	0,00	0,00
CSH	0,00	81,03	4,28	0,00
ANTE	0,00	0,00	3,08	100,00
AR	0,00	0,00	0,00	0,00
ARH	0,00	0,00	60,61	0,00
ARMA	0,00	0,00	2,12	0,00
FA	0,00	5,72	0,00	0,00
HF	0,00	2,98	0,00	0,00
TOEP	100,00	0,00	0,00	0,00
TOEPH	0,00	8,54	15,71	0,00

CS: Simetria Composta; CS* Simetria Composta com transformação para heterogeneidades de variância (h_i/h_m); CSH: Simetria Composta com Variâncias Heterogêneas; ANTE: Antependência (SAD); AR: Autorregressiva de 1ª ordem; ARH: Autorregressiva de 1ª ordem com variâncias heterogêneas; ARMA: Médias Movéis Autorregressiva; FA: Fator Analítica com um termo multiplicativo; HF: Huynh-Feldt; TOEP: Toeplitz; TOEPH: Toeplitz com Variâncias heterogêneas.

Após identificar as estruturas da matriz de covariância R mais condizentes com os dados analisados, procurou-se a estrutura que melhor representasse a covariância associada à interação progênie com colheitas (PC). Para colheitas anuais, a estrutura mais indicada com 95,68% de probabilidade foi a TOEP, indicando que o padrão de interação difere entre as colheitas avaliadas. Para biênios ocorreu exata concordância entre CSH e ARH com 49,99% de chance das mesmas serem as corretas. Para os anos bons, a estrutura

provável com 90,36% de chance foi a ANTE. Quando se analisou a matriz de covariância de PC para os anos ruins verificou-se que a ausência de correlação (VC) foi mais adequada (Tabela 7).

Tabela 7 Probabilidade (W^2_{10}), em porcentagem, de o modelo da estrutura de covariância da interação progênies com colheitas (PC) ser o correto a partir da melhor estrutura residual escolhida na Tabela 6, nos diferentes modos de se considerar os dados

MODELO	ANUAL	BIÊNIO	ANOS BONS	ANOS RUINS
VC	0,00	0,00	0,00	51,29
CS	0,00	0,00	0,00	11,44
CSH	0,00	49,95	4,97	0,00
ANTE	0,00	0,00	90,36	0,00
AR	3,52	0,09	0,00	29,59
ARH	0,00	49,95	0,00	7,67
FA	0,00	0,00	4,49	0,00
TOEP	95,68	0,00	0,00	0,00

VC: componente de variância; CS: Simetria Composta; CSH: Simetria Composta com Variâncias Heterogêneas; ANTE: Antedependência; AR: Autorregressiva de 1ª ordem; ARH: Autorregressiva de 1ª ordem com Variâncias Heterogêneas; FA: Fator Analítica com um termo multiplicativo; TOEP: Toeplitz;

4.3 Consequências na seleção

As consequências das diferentes modelagens e modo de se agrupar os dados foram analisadas de duas maneiras. Primeiro verificou-se a estimativa dos parâmetros de maior importância no melhoramento genético, que são a herdabilidade e o ganho com a seleção. Por último, analisou-se a alteração no ordenamento das progênies quanto a seu valor genético, por meio da correlação classificatória e do índice de coincidência ao se selecionar as melhores

progênies. Adicionalmente, são apresentados os resultados com a alternativa baseada na previa transformação dos dados e posterior análise conjunta admitindo-se a matriz R do tipo simetria composta, aqui denotada por CS*, conforme Resende (2007).

4.3.1 Estimativa dos parâmetros

As estimativas dos parâmetros genéticos variaram entre os modelos com as diferentes matrizes de covariância dos fatores aleatórios assumidas nas análises envolvendo as colheitas individuais (Tabela 8). Destacadamente a análise em parcela subdividida no tempo, CS(R) VC(PC) resultou em menores magnitudes da herdabilidade (0,58) e ganho com a seleção (4,66). A melhor modelagem de covariância escolhida TOEP (R) TOEP (PC) resultou em ganho 21% superior a CS(R) VC (PC), no entanto não apresentou diferença relativa a CS*.

Tabela 8 Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênies das modelagens escolhidas para colheitas anuais

Modelo		Parâmetros			
R	PC	\hat{h}^2	\hat{h}	GS	GS_m
CS	VC	0,58	0,76	4,66 (100%)	11,17%
CS*	VC	0,69	0,83	5,73 (122%)	13,74%
TOEP	VC	0,66	0,81	5,14 (110%)	12,32%
TOEP	TOEP	0,63	0,80	5,64 (121%)	13,52%

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênies com colheitas; VC: componente de variância; CS: Simetria Composta; CS*:

Simetria Composta com transformação para heterogeneidades de variância (\hat{h}_m); TOEP:

Toeplitz; \hat{h}^2 : Herdabilidade; \hat{h} Acurácia. GS : Ganho com a seleção das cinco melhores progênes (Ganho relativo ao padrão CS, em %); GS_m : Ganho com a seleção das cinco melhores progênes.

Em biênios, os resultados obtidos pelas duas melhores modelagens (CSH(R) CSH(PC) e CSH(R) ARH(PC)) mostraram-se equivalentes ao alcançado pelos demais modelos. O ganho genético com essas modelagens foi apenas 5% superior à obtida pelo modelo CS(R) VC (PC), e apenas 3% superior ao modelo CS*. Esse resultado demonstra que ao se agrupar os anos em biênios minimiza-se a questão da heterocedasticidade e promove-se um ajuste no padrão de correlações entre medições (Tabela 9).

Tabela 9 Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênes das modelagens escolhidas para biênios

Modelo		Parâmetros			
R	PC	\hat{h}^2	\hat{h}	GS	GS_m
CS	VC	0,66	0,81	5,34 (100%)	12,80%
CS*	VC	0,69	0,83	5,47 (102%)	13,12%
CSH	VC	0,78	0,88	5,73 (107%)	13,74%
CSH	CSH	0,66	0,81	5,65 (105%)	13,55%
CSH	ARH	0,65	0,80	5,64 (105%)	13,52%

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênes com colheitas; VC: componente de variância; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância (\hat{h}_i/\hat{h}_m); CSH: Simetria Composta com Variâncias Heterogêneas; ARH: Autorregressiva de 1ª ordem com Variâncias Heterogêneas; \hat{h}^2 : Herdabilidade; \hat{h} Acurácia. GS : Ganho com a seleção das cinco melhores progênes (Ganho relativo ao padrão CS em %); GS_m : Ganho com a seleção das cinco melhores progênes, em porcentagem.

A modelagem indicada para as matrizes R e PC para os anos bons resultou em estimativas inferiores da herdabilidade e ganho com a seleção. Vale

ressaltar que o teste de Mauchly havia decidido pela aceitação da hipótese de esfericidade. O modelo CS(R) VC(PC) resultou em melhores estimativas para esses parâmetros, sendo bastante similar ao modelo CS* (Tabela 10).

Tabela 10 Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênes das modelagens escolhidas para os anos bons

Modelo		Parâmetros			
R	PC	\hat{h}^2	\hat{h}	GS	GS _m
CS	VC	0,63	0,79	10,66 (100%)	14,20%
CS*	VC	0,55	0,74	10,91 (102%)	14,54%
ARH	VC	0,17	0,41	3,33 (31%)	4,40%
ARH	ANTE	0,49	0,70	10,59 (0,99%)	14,11%

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênes com colheitas; VC: componente de variância; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância (\hat{h}_i/\hat{h}_m); ARH: Autorregressiva de 1ª ordem com Variâncias Heterogêneas; ANTE: Antedependência; \hat{h}^2 : Herdabilidade; \hat{h} Acurácia. GS: Ganho com a seleção das cinco melhores progênes (Ganho relativo ao padrão CS em %); GS_m: Ganho com a seleção das cinco melhores progênes, em porcentagem.

Em anos ruins, a modelagem ideal aos dados ANTE (R) resultou nas menores estimativas. Caso fosse se modelar apenas por CS (R) ou CS* poderia se pensar que houve ganho nesse caso, principalmente quando se analisa o GS em relação à média original, que foi de 25,26% para CS(R) e 27,90% para CS*. Entretanto, a realidade para esse modo de se considerar os dados é que a herdabilidade e o GS em relação à média antes da seleção são nulos. Esse fato é corroborado pela significância do componente de variância de progênes que foi de 0,11 quando se empregou a CS(R) e de 0,39 para ANTE (R). Para a melhor modelagem o componente da interação progênes com colheitas também se

alterou em relação a CS(R), uma vez que o mesmo foi de 0,06 (CS(R)) para 0,38 em ANTE (R) (Tabela 11).

Tabela 11 Estimativas de herdabilidade (h^2), acurácia (h) e ganho com a seleção (GS) das cinco melhores progênes das modelagens escolhidas para os anos ruins

Modelo		Parâmetros			
R	PC	\hat{h}^2	\hat{h}	GS	GS _m
CS	VC	0,40	0,63	2,14 (100%)	25,26%
CS*	VC	0,47	0,68	2,33 (108%)	27,90%
ANTE	VC	0,00	0,00	0,37 (17%)	0,04%

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênes com colheitas; VC: Ausência de covariância; CS: Simetria Composta (Subdividida); CS*: Simetria Composta (Subdividida) com transformação para heterogeneidades de variância (\hat{h}_i/\hat{h}_m); ANTE: Antedependência; \hat{h}^2 : Herdabilidade; \hat{h} : Acurácia. GS: Ganho com a seleção das cinco melhores progênes (Ganho relativo ao padrão CS em %); GS_m: Ganho com a seleção das cinco melhores progênes, em porcentagem.

Percebeu-se pelas Tabelas 9 a 11 que os diferentes modos de se considerar os dados, aplicando-se as modelagens escolhidas, não diferiram em demasia quanto ao GS em relação à média original da população, exceto para a análise dos anos ruins, apesar de ter havido superioridade quando se utilizou apenas os anos bons na análise. Para colheitas anuais o GS em porcentagem em relação a média foi de 13,52, para biênios foi de 13,55 e para anos bons de 14,20.

4.3.2 Alteração no ordenamento

A alteração no ordenamento das progênies, a qual afeta diretamente o processo seletivo foi demonstrado de duas maneiras. Primeiramente, estimou-se as correlações de Spearman entre as modelagens testadas e depois estimou-se o índice de coincidência ao se selecionar as cinco melhores progênies, uma vez que o interesse dos melhoristas centra-se nesse aspecto.

Em colheitas anuais observou-se elevada correlação no ordenamento dos E-BLUPs a partir dos diferentes modelos. O menor valor de 0,9792 foi obtido entre CS* e TOEP(R). Todas as correlações envolvendo o melhor modelo TOEP(R)TOEP(PC) foram maiores que 0,987. Correlação de 0,9987 foi obtida entre CS(R) e TOEP(R). Elevada correlação também foi obtida entre o ordenamento realizado por TOEP(R) e TOEP(R)TOEP(PC), que foi de 0,9883 (Tabela 12).

Tabela 12 Correlação de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas anuais

	CS*	TOEP(R)	TOEP(R)TOEP(PC)
CS(R)	0,9818**	0,9987**	0,9896**
CS*		0,9792**	0,9870**
TOEP(R)			0,9883**

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênies com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância (\hat{h}_i/\hat{h}_m); TOEP: Toeplitz; ** Correlação Significativa a 5%.

Ao se analisar o índice de coincidência (IC) obtido pela seleção nos diferentes modelos observa-se que houve idêntica seleção entre TOEP(R) e CS(R) e entre TOEP (R)TOEP(PC) com CS*. Os índices nos demais casos foram de 74%, havendo alteração de apenas uma progênie ao se selecionar pelos diferentes modelos (Tabela 13).

Tabela 13 Índice de coincidência de Hamblin e Zimmerman (%) na seleção das cinco melhores progênies por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas anuais

	CS*	TOEP(R)	TOEP(R)TOEP(PC)
CS(R)	74	100	74
CS*		74	100
TOEP(R)			74

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênies com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância ($\hat{\mu}_i/\hat{h}_m$); TOEP: Toeplitz.

Em biênios, houve maior alteração no ordenamento dos E-BLUP's das progênies. Correlação classificatória de aproximadamente 0,70 foi obtida entre as melhores modelagens escolhidas (CSH (R) CSH (PC); CSH (R) ARH (PC)) e a CS(R). A correlação entre CS* e as modelagens indicadas como superiores foram próximas a 0,75. Correlação intermediária foi obtida pelos ordenamentos realizados pelas melhores modelagens (CSH (R) CSH (PC); CSH (R) ARH (PC)) com a abordagem que modelou apenas R como CSH, cujo valor foi de 0,84. Correlações maiores que 0,95 foram obtidas entre os ordenamentos de CS(R), CS* e CSH(R), indicando semelhança entre as análises. (Tabela 14).

Tabela 14 Correlações de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas agrupadas em biênios

	CS*	CSH(R)	CSH(R)CSH(PC)	CSH(R)ARH(PC)
CS(R)	0,9922**	0,9480**	0,6909**	0,7026**
CS*		0,9597**	0,7545**	0,7636**
CSH(R)			0,8441*	0,8480**
CSH(R)CSH(PC)				0,9948**

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênes com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância (\hat{h}_i/\hat{h}_m); CSH: Simetria Composta com Variâncias Heterogêneas; ARH: Autorregressiva de 1ª ordem com Variâncias Heterogêneas; ** Correlação Significativa a 5%.

Corroborando os resultados apresentados na Tabela 14, em que se obtiveram baixas correlações entre as duas melhores modelagens ((CSH (R) CSH (PC) e CSH (R) ARH (PC)) com CS(R) e CS* os IC nesses casos foi de 22%, demonstrando a não coincidência na seleção por esses modelos. As abordagens CS(R), CS* e CSH(R), que obtiveram estimativa de correlação acima de 0,95 (Tabela 14) coincidiram em 74% na seleção das melhores progênes. Baixas coincidências foram obtidas entre a melhor modelagem para R (CSH) quando se compara com CSH (R) CSH (PC), cujo valor de IC foi de 48%, que foi menor ainda quando de utilizou ARH em PC (IC=22%) (Tabela 15).

Tabela 15 Índice de coincidência de Hamblim e Zimmerman (%) na seleção das cinco melhores progênies por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas agrupadas em biênios

	CS*	CSH(R)	CSH(R)CSH(PC)	CSH(R)ARH(PC)
CS(R)	74	74	22	22
CS*		74	22	22
CSH(R)			48	22
CSH(R)CSH(PC)				100

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênies com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância ($\hat{\sigma}_i/\hat{\sigma}_m$); CSH: Simetria Composta com Variâncias Heterogêneas; ARH: Autorregressiva de 1ª ordem com Variâncias Heterogêneas.

Em anos bons, no qual aceitou-se estatisticamente a esfericidade, houve correlação próxima da unidade entre CS(R) e CS*, consideradas as modelagens mais adequadas. Percebe-se que as correlações inferiores foram obtidas quando modelou-se a matriz R sem necessidade, bem como a PC. Apesar de que a modelagem de PC, como apresentado na tabela 10, fez com que essa abordagem se aproximasse da indicada como a melhor (CS(R)) (Tabela 16).

Tabela 16 Correlação de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra alta (Anos Bons)

	CS*	ARH(R)	ARH(R)ANTE(PC)
CS(R)	0,9961**	0,9064**	0,9519**
CS*		0,9039**	0,9610**
ARH(R)			0,8649**

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênies com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância ($\hat{\sigma}_i/\hat{\sigma}_m$); ARH: Autorregressiva de

1ª ordem com Variâncias Heterogêneas; ANTE: Antedependência Estruturada; ** Correlação Significativa a 5%.

Quanto ao IC para os anos bons, perceberam-se resultados um pouco mais drásticos do que na correlação classificatória. Considerando tanto CS(R) como CS*, as quais coincidiram em 100% na seleção realizada, como referência constata-se que os IC foram de 74% em relação à ARH(R) e de 48% quando comparado com ARH (R) ANTE(PC). Interessante observar que o IC foi maior na comparação com ARH (R) apesar deste modelo ter obtido menor correlação classificatória ao compará-lo com CS(R) e CS*, já com ARH(R)ANTE(PC) o IC foi menor, mesmo ele tendo maior correlação com as modelagens ideais (Tabela 17).

Tabela 17 Índice de coincidência de Hamblim e Zimmerman (%) na seleção das cinco melhores progênies por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra alta (Anos Bons)

	CS*	ARH(R)	ARH(R)ANTE(PC)
CS(R)	100	74	48
CS*		74	48
ARH(R)			48

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênies com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância ($\hat{\sigma}_i/\hat{\sigma}_m$); ARH: Autorregressiva de 1ª ordem com Variâncias Heterogêneas; ANTE: Antedependência Estruturada; Índice de coincidência (IC) (HAMBLIN; ZIMMERMAN, 1986).

Em anos ruins baixa correlação entre o ordenamento E-BLUP's foram obtidas entre CS(R) e CS* quando se comparou com a modelagem ideal ANTE (R). O valor estimado foi de 0,7441 em relação à CS e de 0,6909 com CS* (Tabela 18). A baixa correlação obtida refletiu-se no IC, nos quais houve coincidência na seleção de 74% para CS(R) e 48% para CS* quando comparou-se com a estrutura escolhida para essa base de dados (ANTE (R)) (Tabela 19).

Tabela 18 Correlação de Spearman dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra baixa (Anos Ruins)

	CS*	ANTE (R)
CS(R)	0,9896**	0,7441**
CS*		0,6909**

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênes com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância (\hat{h}_i/\hat{h}_m); ANTE: Antedependência Estruturada; ** Correlação Significativa a 5%.

Tabela 19 Índice de coincidência de Hamblim e Zimmerman (%) na seleção das cinco melhores progênes por meio dos E-BLUP's obtidos por diferentes modelagens para colheitas de safra baixa (Anos Ruins)

	CS*	ANTE (R)
CS(R)	74	74
CS*		48

R: Matriz de Covariância residual; PC: Matriz de Covariância do fator de interação progênes com colheitas; CS: Simetria Composta; CS*: Simetria Composta com transformação para heterogeneidades de variância (\hat{h}_i/\hat{h}_m); ANTE: Antedependência Estruturada.

Ao se comparar os modos de se agrupar os dados, visando a melhor alternativa para se proceder à seleção, percebeu-se alta correlação entre colheitas anuais e os anos bons, que foi de 0,9350. Quando se compararam colheitas anuais e os biênios estimou-se uma correlação entre o ordenamento dos E-BLUP's de 0,7259. A correlação entre os biênios e os anos bons foi semelhante à obtida entre colheitas anuais e biênios, uma vez que anos bons foram altamente correlacionados com as colheitas anuais. Correlações não significativas foram observadas em todos os casos em que os anos ruins estavam envolvidos (Tabela 20).

O IC obtido pelos diferentes modos de se agrupar os dados de produção de grãos suporta os resultados encontrados nas correlações acima. IC de 100% foi obtido entre colheitas anuais e anos bons e de 48% considerando-se a seleção realizada por biênios quando comparou-se com colheitas anuais e anos bons. IC nulo foi estimado em todas ocasiões em que a análise conjunta dos anos ruins foi considerada (Tabela 21).

Tabela 20 Correlação de Spearman entre os modos de se agrupar os dados para seleção

	Biênio	Anos Bons	Anos Ruins
Anual	0,7259**	0,9350**	0,0727
Biênio		0,7610**	-0,0766
Anos Bons			-0,1558

*O ordenamento foi obtido por meio da melhor modelagem para cada situação.

** Correlação Significativa a 5%.

Tabela 21 Índice de coincidência (%) na seleção das cinco melhores progênies em cada modo de se agrupar os dados para seleção

	Biênio	Anos Bons	Anos Ruins
Anual	48	100	0
Biênio		48	0
Anos Bons			0

*O ordenamento foi obtido por meio da melhor modelagem para cada situação. Índice de coincidência (IC) (HAMBLIN; ZIMMERMAN, 1986)

5 DISCUSSÃO

Apesar de uma lavoura de café ter vida útil por décadas, a seleção durante as primeiras colheitas foi demonstrada ser eficiente (SERA, 1987). Para se ter maior confiabilidade na seleção de genótipos superiores, a produção necessita ser avaliada por pelo menos quatro colheitas. No presente trabalho foram utilizadas oito avaliações objetivando-se melhor descrever o comportamento produtivo das progênes avaliadas.

Uma condição necessária para que se consiga o progresso genético é que haja suficiente variabilidade genética na população de melhoramento. Por meio das análises por colheita (Tabela 2) e análises conjuntas (Tabelas 8 a 11) evidenciou-se que a herdabilidade é suficiente para auferir-se ganhos com a seleção. Vale ressaltar que nos anos de baixa produtividade ocorreu um incremento da variância ambiental e, por conseguinte, reduzidas estimativas de herdabilidade. Brancourt-Hulmel et al. (2005) observaram esse comportamento ao avaliar progênes de trigo em ambientes contrastantes quanto ao nível de nitrogênio e aplicação ou não de fungicida. Os autores relataram que altas herdabilidades estiveram associadas com maior qualidade ambiental promovida pelo fornecimento de nitrogênio. Entretanto, Presterl et al. (2003) avaliando genótipos de milho quanto a produção de grãos atestaram que as herdabilidades estimadas em ambientes com e sem aplicação de nitrogênio foram semelhantes, apesar das diferentes médias fenotípicas. Para o cafeeiro as herdabilidades médias encontradas nesse trabalho mostraram-se concordantes ao reportado na literatura. Carvalho et al. (2012) relataram que a média das herdabilidades em nível de progênes estimadas em 123 trabalhos foi de 0,51.

Em virtude da bienalidade do cafeeiro, uma estratégia comumente adotada para a seleção constitui-se em agrupar os anos em biênios. Com esse

agrupamento, percebeu-se uma homogeneização das variâncias ao longos dos biênios. Isso ocorreu pois a variabilidade é reduzida uma vez que a bienalidade da produção foi constatada a partir da segunda colheita (Figura 1). Pode-se notar que a média das estimativas da herdabilidade foram maiores em biênios e nos anos bons (Tabela 2). Mendes (1994) relatou que ao considerar os biênios, obteve menores estimativas dos coeficientes de variação ambiental e maiores coeficientes de determinação genética. O autor atribuiu a maior variação genética nesses casos em função da redução da interação progênes com colheitas.

As correlações de Spearman dos E-BLUP's das colheitas individuais e em biênios (Tabelas 3 e 4) nos remetem principalmente a duas constatações. Primeiro é que a interação progênes com colheitas é expressiva no cafeeiro (BOTELHO et al., 2010). A segunda informação importante é que a seleção em colheitas iniciais do cafeeiro é viável, como foi primeiramente demonstrado por Sera (1987). Posteriormente a este trabalho outros autores abordaram a questão, tais como Carvalho et al. (2006), Fazuoli et al. (2005), Mendes (1994) e Oliveira et al. (2011). Atualmente é consenso geral dos melhoristas de café que a seleção pode ser realizada com segurança na quarta colheita, ou na primeira grande safra. Essa afirmação é suportada pelos resultados da Tabela 3, pois a correlação dos E-BLUP's dos anos bons foi elevada a partir da quarta colheita, sendo essa de 0,92 com a última avaliação. Com os dados agrupados em biênios o resultado não foi diferente, obtendo-se o segundo biênio correlação de 0,84 com o último (Tabela 4).

Como a seleção deve ser praticada pela análise conjunta dos dados consideraram-se quatro modos de se agrupá-los nesta análise (Anual, Biênios, Anos Bons e Anos Ruins). Ao se proceder a análise conjunta das colheitas a condição longitudinal dos dados surge no contexto. Essa condição é gerada pela repetição das medições em uma mesma unidade experimental e tem como

principal consequência a dependência entre elas, que é função das variâncias e covariâncias dos fatores aleatórios do modelo estatístico. Comumente esses dados são analisados pelo esquema de parcelas subdivididas no tempo, no entanto sua eficiência pode ser reduzida devido a não esfericidade da matriz de covariâncias dos erros (matriz R). Isso ocorre porque essa abordagem considera que as variâncias são homogêneas entre as colheitas e que a correlação entre cada par de colheitas é a mesma. No presente trabalho, houve rejeição da hipótese de esfericidade em todas as situações analisadas, exceto nos anos bons (Tabela 5), indicando que a análise no esquema de parcelas subdivididas no tempo não é apropriada. Segundo Resende (2007), a rejeição da esfericidade em plantas perenes parece ser regra. De acordo com Piepho, Buchse e Ritchter (2004) a esfericidade é raramente atingida no caráter produção em plantas perenes e que ela pode ter importantes significados biológicos. Desse modo abordagens alternativas devem ser buscadas para solucionar o problema.

Na busca de predições genéticas fidedignas foram realizadas modelagens utilizando-se diferentes estruturas para a matriz de covariâncias dos erros (matriz R), bem como para a matriz de covariâncias da interação progênies com colheitas (matriz PC). Para as colheitas anuais, a indicação da estrutura TOEP para a R e para a interação PC nos indica uma situação completamente diferente da estrutura de simetria composta. As correlações entre os E-BLUP's obtidos em cada colheita nos indicam claramente o padrão produtivo que foi acomodado pela matriz TOEP. Percebe-se que em ocasiões adjacentes a correlação é inexistente, enquanto que entre anos produtivos essas correlações são mais elevadas e positivas (Tabela 3).

Quando os biênios foram considerados, a estrutura mais provável foi a de simetria composta heterogênea (CSH) tanto para a matriz R quanto para a PC. Isso provavelmente pode ter ocorrido em função da menor complexidade dos dados em relação aos demais. Em trabalhos com *Coffea canephora*, alguns

autores verificaram que a CSH foi a mais adequada (CECON et al., 2008; CILAS; MONTAGNON; BAR-HEN, 2011). Outra constatação é que o agrupamento das colheitas em biênios fez com que a correlação dos efeitos do fator considerado não se alterasse em função do intervalo de tempo. Isso quer dizer que a correlação permaneceu a mesma considerando-se qualquer biênio par a par. Portanto, pode-se supor que agrupar as colheitas em biênios pode não ser a melhor alternativa, uma vez que se alterando a estrutura do caráter podem-se perder informações a respeito do comportamento produtivo das progênies.

A consideração apenas dos anos bons na análise fez com que a esfericidade fosse aceita pelo teste empregado (Tabela 5). Isso nos informa que a variância e covariância entre as colheitas (Anos Bons) foram constantes. A partir dessa informação pode-se aplicar testes de F exatos sob modelo fixo. Pode-se inferir também que as estimativas dependentes de variâncias podem ser consideradas precisas. Apesar disso, caso fosse aplicado o método de seleção de modelos, poderia se pensar que a estrutura ARH para o erro e ANTE para a interação PC fosse a correta para os dados em questão. Essas estruturas assumem que a correlação entre as colheitas diminui à medida que o tempo entre elas aumenta. O que mostrou não ser verdadeiro.

Para os anos ruins identificou-se a ANTE para o erro e a VC para a interação PC. Esse resultado chamou a atenção, pois considerando-se a estrutura CS o componente de variância de progênies foi considerado significativo (p -valor = 0,11), bem como o de interação progênies com colheitas (p -valor = 0,06) (Tabela 22A). Quando se alterou a CS pela estrutura indicada como a mais provável aos dados, a probabilidade de se considerar componentes significativos quando eles não existem se elevou a níveis não aceitáveis, induzindo a conclusão de que na realidade a variação para esses fatores não foi detectada. A não existência de interação foi reforçada pela escolha da estrutura VC para o fator PC, pois essa estrutura admite independência entre os níveis do fator em

consideração. Essa constatação é séria na experimentação e é comumente relatada na literatura como aumento na taxa de erro tipo I (EYDURAN; AKBAS, 2010; KESELMAN; ALGINA; KOWALCHUK, 2001). O melhorista pode, erroneamente, selecionar em uma população não variável geneticamente, impedindo o ganho com a seleção. Ou ainda pode tentar procurar genótipos estáveis e adaptáveis a diversos ambientes quando, na verdade, os mesmos não se interagem.

Para atenuar a heterogeneidade de variância e covariância, Resende (2007) sugere a multiplicação dos dados pela razão da raiz quadrada da herdabilidade obtida na colheita analisada pelo valor médio de todas as herdabilidades, abordagem essa denominada neste trabalho de CS*. Assim, após a transformação dos dados, segundo o autor, a estrutura CS poderia ser efetiva. No entanto, isso não ocorreu uma vez que a superioridade não foi confirmada pelos critérios de seleção de modelos em todos os casos considerados, demonstrando que essa transformação corrige somente para a questão da heterocedasticidade.

Percebeu-se, portanto, que o padrão de covariância em uma análise longitudinal é função do caráter avaliado e da influência ambiental na expressão do mesmo. Dessa forma, a identificação da estrutura mais adequada deve ser realizada em toda análise de dados provenientes de medidas repetidas. Procedendo desse modo, o melhorista pode realizar a seleção com menor chance de erro possível, pois como foi percebido pelos resultados, e será discutido a seguir, as diferentes estruturas de covariância trazem consequências ao processo seletivo, seja pela alteração nas estimativas populacionais ou na alteração do ordenamento das progênes. A alteração nas estimativas dos parâmetros e a realização de diferentes seleções já havia sido relatada em alguns trabalhos (APIOLAZA; GILMOUR; GARRICK, 2000; SMITH et al., 2007). Entretanto, a mensuração do impacto real na seleção não foi encontrada.

O foco do melhorista centra-se no ganho com a seleção e sua importância é atestada uma vez que é com base nesse ganho que são traçadas as estratégias a serem seguidas em um programa de melhoramento. Percebeu-se que as diferentes modelagens resultam em diferentes estimativas de herdabilidade e ganho com a seleção das melhores progênies (Tabelas 8 a 11). As modelagens indicadas como as melhores nem sempre obtiveram as maiores estimativas, apesar de esse fato ter ocorrido na maioria das vezes. Entretanto, o que se busca é a interpretação da realidade biológica contida nos dados.

Outra constatação que surge à luz dos resultados (Tabelas 8 a 11) refere-se ao ganho com a seleção obtido pelos diferentes modos de se agrupar os dados. Exceto para os anos ruins, as estimativas do GS em porcentagem da média original, obtidas pelos modelos indicados, não diferiram em demasia havendo ligeira vantagem para os anos bons. Interessante observar nos anos ruins que caso se aplicasse a estrutura CS ou CS* o ganho com a seleção em relação à média antes da seleção seria o maior entre todos os agrupamentos. Essa constatação foi observada por Oliveira et al. (2011) que, analisando a média de duas colheitas ruins percebeu que o ganho em porcentagem obtido por essa estratégia foi o maior dentre todos os analisados. A explicação dos autores para o ocorrido foi em função da média baixa obtida por esse agrupamento, o que ocorreu também no presente trabalho. O ponto crucial nessa questão é que a estrutura indicada como a provável aos anos ruins fez com que o ganho em porcentagem fosse próximo de zero, como o foi em unidades do caráter. Dessa forma, percebe-se que novamente a estrutura correta aos dados nos informou sobre uma realidade que ficaria desconhecida sob simetria composta. Em Oliveira et al. (2011), caso os autores tivessem utilizado modelagens alternativas poderiam ter encontrado resultado diferente para o GS em relação à média nos anos ruins que, pelo bom resultado, poderia até induzir à prática da seleção nesses anos. No presente trabalho, em termos de ganho seletivo, a recomendação

seria a de se fazer a seleção por meio dos dados obtidos em anos de alta produção.

Apesar do foco do melhorista ser no ganho com a seleção, na prática da seleção o que parece realmente importar é se as progênies selecionadas são realmente superiores quanto ao seu valor genético para a produção. Portanto, outro foco em que o melhorista deve estar sempre atento é no ordenamento das progênies genotipicamente superiores. A partir das diferentes modelagens utilizadas, bem como o modo como se agrupam os dados, houve alteração no ordenamento dos E-BLUP's das progênies e também modificaram quais as progênies seriam selecionadas em cada abordagem (Tabelas 12 a 21). Constatou-se, portanto, a relevância das modelagens obtidas para o melhoramento genético do cafeeiro.

Em colheitas anuais ocorreu um fato interessante quanto a se modelar o erro e também a interação. Quando se modelou R e PC por TOEP houve exata coincidência da seleção realizada quando comparado com CS*, a qual resultou na maior estimativa para o GS (Tabela 8). Percebe-se com isso que a recomendação de Resende (2007) em favor do uso da simetria composta com correção para heterogeneidades de variância foi válida para o melhoramento da produção de grãos no cafeeiro. Entretanto, caso se realizasse apenas a modelagem para a matriz R por TOEP a concordância passou a ser para CS, deixando de ser a correção para a heterogeneidade considerada como adequada. Outro comentário a favor do uso da modelagem para PC foi referente ao GS quando comparado com a abordagem que modelou apenas R como TOEP, sendo essa 11% inferior.

Como comentado anteriormente, existe consenso no uso dos dados de produção do cafeeiro como biênios. Nesse caso, percebeu-se que modelagens alternativas indicadas, em detrimento da CS, promoveram maiores alterações no ordenamento das progênies (Tabelas 14 e 15), apesar de não ter havido

considerável diferença em termos de GS (Tabela 9), pois o mesmo foi de 5% em relação ao padrão CS e de 0,25% em relação à média antes da seleção. Poderia se pensar que isso ocorreu provavelmente devido ao fato de a alteração no ordenamento ter ocorrido com maior intensidade abaixo do ponto de truncamento da seleção, sendo concordantes as progênes selecionadas. Porém a estimativa do IC rejeita essa hipótese, uma vez que os IC também foram baixos. Portanto acredita-se que a alteração no ordenamento tenha ocorrido com progênes próximas quanto a seu desempenho. Novamente, a modelagem de PC, além de R, promoveu recuperação da informação contida nesse fator, uma vez que caso se tivesse modelado apenas a matriz R se observaria elevada concordância entre a seleção realizada comparando-se com CS e também com CS*. Dessa forma, poderia se concluir que modelagens alternativas não seriam efetivas, mesmo na ausência de esfericidade. A semelhança entre CS* e CSH foi relatada por Resende (2007) e é referendada pelos resultados obtidos.

Considerando-se os anos bons percebe-se um padrão interessante que merece ser comentado. Observou-se que a o uso de uma estrutura de covariância em R ao invés de CS, quando a mesma não é necessária, promoveu acentuada queda na estimativa do ganho com a seleção (Tabela 10). Com isso, esperava-se que houvesse maior alteração no ordenamento realizado pelas duas modelagens, o que não ocorreu (Tabelas 16 e 17). Provavelmente a diferença residiu em maior parte nas estimativas dos E-BLUP de cada progênie do que no seu ordenamento. Portanto, as progênes foram igualmente afetadas pela diferenciação na estrutura. Outro ponto que merece destaque refere-se à modelagem da matriz PC. Percebeu-se que o GS obtido por essa abordagem (ARH (R) ATNTE (PC)) foi praticamente o mesmo ao de CS, considerada como sendo a estrutura ideal. Entretanto, o IC entre essas duas modelagens foi de apenas 48%. Neste caso percebeu-se também que mediante o uso somente da correlação para se estimar as diferenças no ordenamento essa diferença não seria

detectada, o que demonstra a importância de se verificar também os IC's. Com relação à dúvida do uso de CS ou CS* que persistia nesse caso, uma vez que CS* promoveu maior GS, foi sanada pelo IC, pois o mesmo foi de 100%.

Em função do exposto até o momento provavelmente as análises realizadas nos anos ruins foram as mais surpreendentes. Nesse caso, como já comentado, a modelagem indicada como ideal forneceu a realidade presente nos dados, evitando-se aceitar erroneamente a hipótese de existência de variação de progênes e de sua interação com as colheitas (Tabela 22A) e demonstrando que o ganho de seleção para esse agrupamento não existe (Tabela 11). Esses comportamentos não foram tão drásticos quanto ao ordenamento E-BLUP uma vez que a seleção realizada por ANTE (R) diferiu em apenas uma progênie comparando-se com CS (IC=74%).

Percebe-se, portanto, os impactos das diferentes estruturas das matrizes de covariância no melhoramento do cafeeiro. Os índices de coincidência entre as modelagens na seleção das progênes atingiram valores que afetam diretamente sua eficiência. Se as progênes selecionadas não se constituírem realmente de genótipos superiores o ganho com a seleção pode ser reduzido. Isso torna-se mais grave quando se trabalha com espécie cujo ciclo de melhoramento é longo, como é o caso do cafeeiro. O melhorista corre o risco de ao final do processo ter a decepção de, após tanto trabalho e recursos investidos, sua cultivar não superar as preexistentes. A alteração no ordenamento das progênes promovida por modelagens mais adequadas aos dados nos aproxima da verdade, pois a busca não é pela complexidade dos modelos, e sim pela verossimilhança com parcimônia. Deve-se lembrar que nunca se deve utilizar um modelo sem verificar sua adequação e quais as consequências nos resultados. Interpretações valiosas podem surgir dos dados e ajudar o melhorista no entendimento da característica para maximizar a eficiência do processo seletivo.

Provavelmente, o resultado obtido de cunho prático de maior importância seja com relação ao modo como se agrupam os dados. Percebeu-se claramente a total coincidência da seleção realizada utilizando-se colheitas anuais e anos bons (Tabelas 20 e 21). A relevância dessa constatação na rotina de um programa de melhoramento para produção de grãos advém da economia de recursos. Podendo-se colher de forma experimental somente nos anos de alta produção, nos anos de baixa, os experimentos podem ser colhidos como lavouras comerciais, economizando-se assim a quantidade de trabalho. Sabe-se que em um programa de melhoramento há diversos experimentos sendo conduzidos ao mesmo tempo, o que demanda considerável força de trabalho para a colheita, que na maioria das vezes é escassa. Além disso, caso o café seja colhido por funcionários contratados, o preço do litro de café colhido pode ser reduzido, impactando diretamente no indicador quantidade de dinheiro gasto por unidade de ganho com a seleção. Outro fato que corrobora a recomendação exposta é que o ganho em porcentagem em relação à média antes da seleção também foi superior para os anos bons.

Oliveira et al. (2011) encontraram resultados semelhantes e afirmaram que a seleção deve ser realizada após a primeira grande safra. Para o presente trabalho isso ocorreria após a segunda colheita, o que devido ao exposto, parece ser plausível. Esses autores também notaram alta correlação entre o ordenamento para as colheitas anuais, média dos anos bons e, semelhante ao observado (Tabelas 20 e 21), baixa coincidência quando as colheitas de baixa produção foram consideradas na análise.

Em função de alguns modelos disponíveis suportarem a heterogeneidade de variâncias faz com que o uso das colheitas anuais seja viável. Desse modo, deve-se evitar agrupar os dados sob qualquer forma, analisando-os como os mesmos são obtidos na prática. Percebeu-se que apesar de os ganhos seletivos ao se comparar a seleção realizada com base nas colheitas anuais e com os biênios

ter sido igual (Tabelas 8 e 9) houve somente 48% de coincidência entre as progênies selecionadas (Tabela 21). Era de se esperar que a coincidência entre as progênies selecionadas por colheitas anuais ou biênios fossem altas. Portanto esses resultados reforçam a discussão se a bienalidade deve ser tratada agrupando-se os dados em biênios. Porém, o agrupamento em biênios é prática comum nas avaliações de produção (BONOMO et al., 2004; BOTELHO et al., 2010; CARVALHO et al., 2006; MENDES, 1994; SERA, 1987). Acredita-se que a análise realizada por meio das colheitas anuais pode evitar a perda de informações a respeito do caráter, pois os biênios alteraram o padrão de variâncias e covariâncias dos dados.

6 CONCLUSÃO

Existem estruturas de covariância dos fatores aleatórios do modelo estatístico que são mais adequadas aos dados de produção de grãos no melhoramento do café arábica do que a simetria composta. As diferentes estruturas variam em função dos dados e sua viabilidade deve ser analisada em cada situação somente quando for requisitada. As diferentes estruturas alteram as estimativas dos parâmetros populacionais, o ordenamento das progênes quanto seu valor reprodutivo e também a seleção realizada por cada abordagem.

O modo como se consideram os dados modifica a seleção das progênes e também a estimativa dos parâmetros. A seleção deve ser praticada em colheitas anuais ou considerando-se apenas os anos de safra alta.

REFERÊNCIAS

AGWANDA, C. O. et al. Selection for bean and liquor qualities within related hybrids of Arabica coffee in multilocal Field trials. **Euphytica**, Wageningen, v. 131, n. 1, p. 1-14, 1993.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Boston, v. 19, n. 6, p. 716-723, Dec. 1974.

ALBERT, P. S. Longitudinal data analysis (repeated measures) in clinical trials. **Statistics in Medicine**, New York, v. 18, n. 13, p. 1707-1732, July 1999.

ALMEIDA, S. R.; CARVALHO, A. Competição de linhagens das variedades comerciais de café arabica Mundo Novo e Catuaí: resultados preliminares. In: CONGRESSO BRASILEIRO DE PESQUISAS CAFEEIRAS, 14., 1987, Campinas. **Resumos...** Rio de Janeiro: IBC, 1986. p. 67-68.

APIOLAZA, L. A.; GARRICK, D. J. Analysis of longitudinal data from progeny tests: some multivariate approaches. **Forest Science**, Bethesda, v. 47, n. 2, p. 129-140, May 2001.

APIOLAZA, L. A.; GILMOUR, A. R.; GARRICK, D. J. Variance modelling of longitudinal height data from a pinus radiata progeny test. **Canadian Journal of Forest Research**, Ottawa, v. 30, n. 4, p. 645-654, 2000.

ARNAU, J. et al. General linear mixed model for analysing longitudinal data in developmental research. **Perceptual and Motor Skills**, Missoula, v. 110, n. 2, p. 547-566, June 2010.

BEARZOTI, E. Emprego de marcadores em modelos mistos. In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 4., 2002, Lavras. **Anais...** Lavras: UFLA, 2002. p. 21-43.

BEGAZO, J. C. E. O.; VIEIRA, J. M.; PAULA, J. F. de. Competição de cultivares de café (*Coffea arabica* L.). In: CONGRESSO BRASILEIRO DE

PESQUISAS CAFEEIRAS, 5., 1977, Guarapari. **Resumos...** Rio de Janeiro: Instituto Brasileiro do Café; GERCA, 1977. p. 142-143.

BERNARDO, R. **Breeding for quantitative traits in plants**. Woodbory: Stemma, 2010. 390 p.

BERTRAND, B. et al. Performance of *Coffea arabica* F1 hybrids in agroforestry and full-sun cropping systems in comparison with American pure line cultivars. **Euphytica**, Wageningen, v. 181, n. 1, p. 147-158, 2011.

BONOMO, P. et al. Avaliação de progênies obtidas de cruzamentos de descendentes do Híbrido de Timor com as cultivares Catuaí Vermelho e Catuaí amarelo. **Bragantia**, Campinas, v. 63, n. 2, p. 207-219, 2004.

BOTELHO, C. E. et al. Adaptabilidade e estabilidade fenotípica de cultivares de café arábica em Minas Gerais. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 12, p. 1404-1411, dez. 2010.

BRANCOURT-HULMEL, M. et al. Indirect versus direct selection of winter wheat for low-input or high-input levels. **Crop Science**, Madison, v. 45, n. 4, p. 1427-1431, July/Aug. 2005.

BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference understanding AIC and BIC in model selection. **Sociological Methods & Research**, Beverly, v. 33, n. 2, p. 261-304, Nov. 2004.

CARVALHO, A. et al. Melhoramento do cafeeiro IV: café Mundo Novo. **Bragantia**, Campinas, v. 12, n. 4/6, p. 97-129, 1952.

_____. Melhoramento do cafeeiro XXII: resultados obtidos no ensaio de seleções regionais de Campinas. **Bragantia**, Campinas, v. 20, p. 711-740, 1961.

CARVALHO, A.; FAZUOLI, L. C. Café. In: FURLANI, A. M. C.; VIEGAS, G. A. (Ed.). **O melhoramento de plantas no Instituto Agrônomo**. Campinas: Instituto Agrônomo, 1993. v. 1, p. 29-76.

CARVALHO, A.; KRUG, C. A. Agentes da polinização flor do cafeeiro *Coffea arabica* L. **Bragantia**, Campinas, v. 9, p. 11-24, 1949.

- CARVALHO, G. R. et al. Comportamento de progênies de cafeeiro cultivar Mundo Novo. **Ciência e Agrotecnologia**, Lavras, v. 30, n. 5, p. 853-860, set./out. 2006.
- CARVALHO, S. P. et al. Meta-análise para estimativas de herdabilidade de caracteres vegetativos e reprodutivos de *Coffea arabica* L. **Semina: Ciências Agrárias**, Londrina, v. 33, n. 4, p. 1291-1298, jul./ago. 2012.
- CARVALHO, W. et al. Comportamento em condições de campo de cafeeiros (*Coffea arabica* L.) propagados vegetativamente e por sementeira. **Coffee Science**, Lavras, v. 3, n. 2, p. 108-114, jul./dez. 2008.
- CECON, P. R. et al. Análise de medidas repetidas na avaliação de clones de café Conilon. **Pesquisa Agropecuária Brasileira**, Brasília, v. 43, n. 9, p. 1171-1176, set. 2008.
- CHENG, J. et al. Real longitudinal data analysis for real people: building a good enough mixed model. **Statistics in Medicine**, New York, v. 29, n. 4, p. 504-520, Feb. 2010.
- CILAS, C. et al. Prediction of genetic value for coffee production in *Coffea arabica* from a half-diallel with lines and hybrids. **Euphytica**, Wageningen, v. 181, n. 1, p. 49-59, 1998.
- CILAS, C.; MONTAGNON, C.; BAR-HEN, A. Yield stability in clones of *Coffea canephora* in the short and medium term: longitudinal data analyses and measures of stability over time. **Tree Genetics and Genome**, Heidelberg, v. 7, n. 2, p. 421-429, Apr. 2011.
- CNAAN, A.; LAIRD, N. M.; SLASOR, P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. **Statistics in Medicine**, New York, v. 16, n. 20, p. 2349-2380, Oct. 1997.
- COLLINS, L. M. Analysis of longitudinal data: the integration of theoretical model, temporal design, and statistical model. **Annual Review of Psychology**, Palo Alto, v. 57, p. 505-529, 2006.

DIGGLE, P. J. et al. **Analysis of longitudinal data**. Norfolk: Oxford University, 2002. 379 p.

EFRON, B.; MORRIS, C. Stein's paradox in statistics. **Scientific American**, New York, v. 236, n. 5, p. 119-127, 1977.

EISENHART, C. The assumptions underlying the analysis of variance. **Biometrics**, Washington, v. 3, n. 1, p. 1-21, 1947.

ETIENNE, H. et al. Biotechnological applications for the improvement of coffee (*Coffea arabica* L.). **In Vitro Cellular & Developmental Biology-Plant**, Montpellier, v. 38, p. 129-138, Mar./Apr. 2002.

EVERITT, B. S. Analysis of longitudinal data: beyond MANOVA. **British Journal of Psychiatry**, London, v. 172, n. 1, p. 7-10, 1999.

EYDURAN, E.; AKBAS, Y. Comparison of different covariance structure used for experimental design with repeated measures. **The Journal of Animal & Plant Science**, Lahore, v. 20, n. 1, p. 44-51, 2010.

FAZUOLI, L. C. et al. Avaliação das cultivares Mundo Novo, Bourbon Amarelo e Bourbon Vermelho de *Coffea arabica* L. em Campinas, SP. **Bragantia**, Campinas, v. 64, n. 4, p. 533-546, 2005.

_____. Cultivares de café selecionadas pelo Instituto Agronômico de Campinas. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 1., 2000, Poços de Caldas. **Anais...** Brasília: EMBRAPA Café; MINASPLAN, 2000. p. 488-493.

GLESER, L. J. A note on the sphericity test. **Annals of Mathematical Statistics**, Ann Arbor, v. 37, n. 2, p. 464-467, 1966.

GONÇALVES, M. C.; FRITSCHÉ-NETO, R. **Tópicos especiais de biometria no melhoramento de plantas**. Viçosa, MG: UFV, 2012. 282 p.

GREENHOUSE, S. W.; GEISSER, S. On methods in the analysis of profile data. **Psychometrika**, Williamsburg, v. 24, n. 2, p. 95-112, 1959.

GURKA, M. J. Selecting the best linear mixed model under REML. **The American Statistician**, Washington, v. 60, n. 1, p. 19-26, 2006.

HAMBLIN, J.; ZIMMERMAN, M. J. O. Breeding common bean for yield mixtures. **Plant Breeding Reviews**, Westport, v. 4, p. 245-272, 1986.

HEDRICK, P. W. **Genetics of populations**. 3rd ed. Boston: Jones and Bartlett, 2005. 737 p.

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

_____. Estimation of genetic parameters. **Annals of Mathematical Statistics**, Ann Arbor, v. 21, p. 309, 1950. Abstract.

HUYNH, H.; FELDT, L. S. Conditions under which mean square ratios in repeated measurement designs have exact F-distributions. **Journal of the American Statistical Association**, New York, v. 65, n. 332, p. 1582-1589, Dec. 1970.

_____. Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. **Journal of Educational Statistics**, Washington, v. 1, n. 1, p. 69-82, Mar. 1976.

KESELMAN, H. J.; ALGINA, J.; KOWALCHUK, R. K. The analysis of repeated measures design: a review. **British Journal of Mathematical and Statistical Psychology**, London, v. 54, n. 1, p. 1-20, May 2001.

KNAFL, G. J.; BEEBER, L.; SCHWARTZ, T. A. A strategy for selecting among alternative models for continuous longitudinal data. **Research in Nursery & Health**, Wageningen, v. 35, n. 6, p. 647-658, Aug. 2012.

KOEHLER, A. B.; MURPHREE, E. S. A comparison of the Akaike and Schwarz criteria for selecting model order. **Journal of the Royal Statistical Society, Series B**, London, v. 37, n. 2, p. 195-197, 1988.

LEHMAN, E. L. Some model I problems of selection. **Annals of Mathematical Statistics**, Ann Arbor, v. 32, n. 4, p. 990-1012, 1961.

LITTELL, R. C. Analysis of unbalanced mixed model data: a case study comparison of ANOVA versus REML/GLS. **Journal of Agricultural, Biological and Environmental Statistics**, Alexandria, v. 7, n. 4, p. 427-491, 2002.

LITTELL, R. C. et al. **SAS for mixed models**. 2nd ed. Cary: SAS Institute, 2006. 817 p.

LITTELL, R. C.; PENDERGAST, J.; RANJINI, N. Modeling covariance structure in the analysis of repeated measures data. **Statistics in Medicine**, New York, v. 19, n. 13, p. 1793-1819, July 2000.

LIU, S.; ROVINE, M. J.; MOLENAAR, C. M. Selecting a linear mixed model for longitudinal data: repeated measures analysis of variance, covariance pattern model, and growth curve approaches. **Psychological Methods**, Washington, v. 17, n. 1, p. 15-30, Jan. 2012.

LUSH, J. L. Family merit and individual merit as bases for selection. **American Naturalist**, Chicago, v. 81, n. 799, p. 241-379, July/Aug. 1947.

_____. The number of daughters necessary to prove a sire. **Journal of Dairy Science**, Champaign, v. 14, n. 3, p. 209-220, May 1931.

MARIGUELE, K. H. et al. Métodos de análise de dados longitudinais para o melhoramento genético da pinha. **Pesquisa Agropecuária Brasileira**, Brasília, v. 46, n. 12, p. 1657-1664, dez. 2011.

MAUCHLY, J. W. Significance test for sphericity of n-variate normal population. **Annals of Mathematical Statistics**, Ann Arbor, v. 11, p. 204-209, 1940.

MEDINA FILHO, P. H.; BORDIGNON, R.; CARVALHO, C. H. S. Desenvolvimento de novas cultivares de café arábica. In: CARVALHO, C. H. S. (Ed.). **Cultivares de café: origem características e recomendações**. Brasília: EMBRAPA Café, 2008. p. 79-102.

MENDES, A. N. G. **Avaliação de metodologias empregadas na seleção de progênies do cafeeiro (*Coffea arabica* L.) no Estado de Minas Gerais.** 1994. 167 f. Tese (Doutorado em Fitotecnia) - Universidade Federal de Lavras, Lavras, 1994.

NAKAGAWA, S.; SCHIELZETH, H. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. **Biological Reviews**, Cambridge, v. 85, n. 4, p. 935-956, Nov. 2010.

OLIVEIRA, A. C. B. et al. Prediction of genetic gains from selection in Arabica coffee progenies. **Crop Breeding and Applied Biotechnology**, Londrina, v. 11, n. 2, p. 106-113, June 2011.

PETEK, M. R.; SERA, T.; FONSECA, I. C. B. de. Predição de valores genéticos aditivos na seleção visando obter cultivares de café mais resistentes à ferrugem. **Bragantia**, Campinas, v. 67, n. 1, p. 133-140, 2008.

PIEPHO, H. P.; BUSCHE, A.; EMRICH, A. Hitchhiker's guide to mixed models for randomized experiments. **Journal of Agronomy and Crop Science**, Hoboken, v. 189, n. 5, p. 310-322, Sept. 2003.

PIEPHO, H. P.; BUCHSE, A.; RICHTER, C. A mixed modelling approach for randomized experiments with repeated measures. **Journal of Agronomy and Crop Science**, Hoboken, v. 190, n. 4, p. 230-247, Aug. 2004.

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, Wageningen, v. 161, n. 1/2, p. 209-228, May 2008.

PLETCHER, S. D.; GEYER, C. J. The genetic analysis of age-dependent traits: modeling a character process. **Genetics**, Austin, v. 153, n. 2, p. 825-833, Oct. 1999.

POL, M. van de; VERHULST, S. Age-dependent traits: a new statistical model to separate within- and between-individual effects. **The American Naturalist**, Chicago, v. 167, n. 5, p. 764-771, May 2006.

PRESTERL, T. et al. Improvement nitrogen use efficiency in European maize: estimation of quantitative genetic parameters. **Crop Science**, Madison, v. 43, n. 4, p. 1259-1265, July/Aug. 2003.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, 2008. Software.

RAMALHO, M. A. P.; CARVALHO, B. L.; NUNES, J. A. R. Perspective for the use of quantitative genetics in breeding of autogamous plants. **ISRN Genetics**, Cairo, v. 2013, p. 1-6, 2013.

RAMALHO, M. A. P. et al. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Lavras: UFLA, 2012. 522 p.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. de. **Experimentação em genética e melhoramento de plantas**. 3. ed. Lavras: UFLA, 2012. 305 p.

RESENDE, M. D. V. de. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: EMBRAPA Florestas, 2007. 561 p.

RESENDE, M. D. V. de; DIAS, L. A. S. Aplicação da metodologia de modelos mistos (REML/BLUP) na estimação de parâmetros genéticos e predição de valores genéticos em espécies frutíferas. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 22, n. 1, p. 44-52, 2000.

RESENDE, M. D. V. de; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, Goiânia, v. 37, n. 3, p. 182-194, set. 2007.

RESENDE, M. D. V. de et al. Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. **Bragantia**, Campinas, v. 60, n. 3, p. 185-193, 2001.

RESENDE, M. D. V. de; STURION, J. A. Spatial statistical analysis of trials through individual mixed models with modelling of errors by ARIMA processes in two dimensions. **Journal of Mathematical & Statistics**, Adelaide, v. 21, n. 1, p. 7-33, Jan. 2003.

ROCHA, T. R.; CARVALHO, A.; FAZUOLI, L. C. Melhoramento do cafeeiro XXXVIII: observações sobre progênies do cultivar Mundo Novo de Coffea arabica na estação experimental de Mococa. **Bragantia**, Campinas, v. 39, n. 15, p. 147-160, 1980.

SAKIYAMA, N. S. et al. Melhoramento do café arábica. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa, MG: UFV, 2005. p. 203-223.

SCHWARZ, G. Estimating the dimensional of a model. **The Annals of Statistics**, Hayward, v. 6, n. 2, p. 461-464, Mar. 1978.

SEARLE, S. R. "C. R. HENDERSON", the statistician and his Contributions to Variance Components Estimation. **Journal of Dairy Science**, Champaign, v. 74, n. 11, p. 4035-4044, Nov. 1991.

SERA, T. Coffee genetic breeding at IAPAR. **Crop Breeding and Applied Biotechnology**, Londrina, v. 1, n. 2, p. 179-199, 2001.

_____. **Possibilidade de emprego de seleção nas colheitas iniciais de café (Coffea arabica L.. cv. Acaiá)**. 1987. 147 f. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 1987.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **Journal of Agricultural Science**, Cambridge, v. 143, n. 6, p. 449-462, Sept. 2005.

SMITH, A. B. et al. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. **Euphytica**, Wageningen, v. 157, n. 1/2, p. 253-266, Sept. 2007.

STATISTICAL ANALYSIS SYSTEM INSTITUTE. **SAS/STAT 9.0: user' guide**. Cary, 2009. Software.

STEEL, R. G. D.; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics: a biometrical approach**. 3rd ed. New York: McGraw Hill, 1997. 666 p.

VALLEJO, G. et al. Selecting the best unbalanced repeated measures model. **Behavior Research Methods**, Austin, v. 43, n. 1, p. 18-36, Mar. 2011.

WALYARO, D. J. A. **Considerations in breeding for improved yield and quality in arabica coffee (*Coffea arabica* L.)**. 1983. 119 f. Thesis (Ph.D. in Genetics and Plant Breeding) - University of Wageningen, Wageningen, 1983.

WHITE, T. L.; HODGE, G. R. Best linear prediction of breeding values in a forest tree improvement program. **Theoretical Applied Genetics**, Berlin, v. 76, p. 719-727, 1988.

APÊNDICE

Conforme mencionado no material e métodos a modelagem do resíduo e da interação progênes com colheitas foi realizada de forma seqüencial. Primeiramente procedeu-se a análise por meio da rotina 1 com cada estrutura

ROTINA 1

MODELAGEM DA MATRIZ DE COVARIÂNCIA DO ERRO (R)

DATA PRODUCAO;

INPUT GEN MED REP Y1;

DATALINES;

“Base de dados”

PROC PRINT;RUN;

%MACRO E_SIGMA(E,EST,S,VAL);

PROC MIXED DATA=&E COVTEST;

CLASS GEN MED REP;

MODEL Y1= REP|MED;

RANDOM GEN GEN*MED;

REPEATED /SUB=REP*GEN TYPE=&EST R RCORR;

MAKE "FitStatistics" OUT=&S(RENAME=(VALUE=&VAL));

RUN;QUIT;

%MEND E_SIGMA;

****ESTRUTURAS DA MATRIZ DE COVARIÂNCIA DO ERRO****

Após executar a macro acima deve-se realizar a análise com cada uma das estruturas de covariância do erro abaixo e anotar o valor da estimativa BIC que se encontra na saída do SAS ®.

```
%E_SIGMA(A,VC,SDS_VC,V_VC);
%E_SIGMA(A,CS,SDS_CS,V_CS);
%E_SIGMA(A,CSH,SDS_CSH,V_CSH);
%E_SIGMA(A,HF,SDS_HF,V_HF);
%E_SIGMA(A,ANTE(1),SDS_ANTE,V_ANTE);
%E_SIGMA(A,AR(1),SDS_AR,V_AR);
%E_SIGMA(A,ARH(1),SDS_ARH,V_ARH);
%E_SIGMA(A,ARMA(1,1),SDS_ARMA,V_ARMA);
%E_SIGMA(A,FA(1),SDS_FA1,V_FA1);
%E_SIGMA(A,TOEP,SDS_TOEP,V_TOEP);
%E_SIGMA(A,TOEPH,SDS_TOEPH,V_TOEPH);
%E_SIGMA(A,UN,SDS_UN,V_UN);
%E_SIGMA(A,UNR,SDS_UNR,V_UNR);
```

ROTINA 2

Perante a identificação da estrutura de covariância para o erro mais condizente com o caráter deve-se modelar também a matriz de covariância do fator de interação progenies com colheitas (PC). Essa modelagem é obtida especificando-se para o erro a estrutura escolhida como mais provável por meio de BIC (menor valor) e se alterar, uma a uma, as estruturas de covariância que

serão testadas na matriz PC. As estruturas testadas são as mesmas que foram utilizadas na modelagem de R e são especificadas por meio do comando TYPE. Este comando especifica a estrutura para a matriz R em REPEATED e em RANDOM o TYPE indica a estrutura utilizada em PC. O menor valor de BIC após este procedimento demonstrará qual a estrutura de covariância que melhor se ajusta aos dados em questão.

```
PROC MIXED DATA=A COVTEST;
CLASS MED GEN REP;
MODEL Y1= REP|MED;
RANDOM MED/ SUB=GEN TYPE=TOEP solution; * E-BLUP das progênies
em cada colheita;
REPEATED MED /SUB=rep*gen TYPE=TOEP;
RUN;QUIT;
```

Tabela 22: Componentes de variância para cada modo de se considerar os dados na análise em esquema de parcelas subdividida no tempo.

Análise	p-valor fator progênie	p-valor fator interação PC
Colheitas Anuais	0,0361	<,0001
Biênio	0,0195	0,0004
Anos Bons	0,0260	<,0001
Anos Ruins	0,1195	0,0697