

KENY HENRIQUE MARIGUELE

**MÉTODOS DE ANÁLISE DE DADOS LONGITUDINAIS NO  
MELHORAMENTO GENÉTICO DA PINHA (*ANNONA SQUAMOSA* L.)**

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

VIÇOSA  
MINAS GERAIS - BRASIL  
2010

**KENY HENRIQUE MARIGUELE**

**MÉTODOS DE ANÁLISE DE DADOS LONGITUDINAIS NO  
MELHORAMENTO GENÉTICO DA PINHA (*ANNONA SQUAMOSA* L.)**

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

APROVADA: 31 de agosto de 2010.

---

Prof. Marcos Deon Vilela de Resende  
(Co-orientador)

---

Prof. Cosme Damião Cruz  
(Co-orientador)

---

Prof. Luiz Antônio dos Santos Dias

---

Prof. Fabyano Fonseca e Silva

---

Prof. José Marcelo Soriano Viana  
(Orientador)

À minha família  
OFEREÇO.

## **AGRADECIMENTOS**

A UFV e, em especial, ao programa de Genética e Melhoramento;

A CAPES e ao CNPQ pela bolsa de estudo;

Ao professor José Marcelo Soriano Viana pela disponibilidade em me orientar desde o primeiro contato e pela agradável convivência;

Um agradecimento especial ao Marcos Deon por em tão pouco tempo aceitar me orientar nesta tese, realmente sem você eu não teria conseguido

Aos professores Cosme, Luiz Antônio e Fabyano por aceitarem, gentilmente, participar da banca de qualificação e defesa;

Aos professores Pedro, Marcos, Karla pelo aprendizado nas disciplinas;

Ao professor da ESAM, Paulo Sérgio Lima e Silva, meu orientador de iniciação científica, que eu tanto admiro e que sempre foi um grande incentivador na minha vida acadêmica.

Ao professor Glauco Miranda, por aceitar participar da banca de qualificação e, posteriormente, dar-me um voto de confiança para trabalhar no Laboratório de Biotecnologia.

Aos queridos amigos Lelis, Mari, Dani, Júlio e Flávia, Viçosa jamais seria a mesma sem vocês, foi um prazer conhecê-las, muito obrigado pela convivência, pelos estudos, pela diversão, pela descontração, enfim, pela amizade; Lelis depois que você foi embora sentimos muito com sua ausência;

Ao Filipe Knop, aluno de IC, foi muito bom conviver com você, parabéns por seu esforço, sua dedicação. Você tem um imenso potencial.

Aos colegas do curso Isis, Ramon, Helder, Renata, Vinícius, Demerson, Loreta, Cássia, Ana Paula e Gilmar;

Às secretárias Rose e Edna por sempre estarem aptas a ajudar os alunos do programa.

**OBRIGADO!**

## SUMÁRIO

RESUMO .....	VI
ABSTRACT.....	VIII
INTRODUÇÃO .....	1
MATERIAL E MÉTODOS .....	5
RESULTADOS E DISCUSSÃO .....	11
CONCLUSÃO .....	15
REFERÊNCIAS BIBLIOGRÁFICAS.....	16
ANEXOS .....	25

## RESUMO

MARIGUELE, Keny Henrique, D.Sc., Universidade Federal de Viçosa, Agosto, 2010.

**Métodos de análise de dados longitudinais no melhoramento genético da pinha (*Annona squamosa* L.).** Orientador: José Marcelo Soriano Viana. Co-orientadores: Marcos Deon Vilela de Resende e Cosme Damião Cruz.

A pinha ou fruta-do-conde (*Annona squamosa* L.) é uma fruteira importante para o Brasil, especialmente na Região Nordeste. Nessa espécie, os caracteres são avaliados repetidas vezes no decorrer da vida do organismo e são denominados infinitamente dimensionais, no sentido de que, em cada unidade de tempo ou idade, o caráter pode ser avaliado, gerando um conjunto multidimensional de dados. O interesse na análise desse tipo de dados geralmente reside na predição de valores dos indivíduos e progênies para determinado ponto no tempo ou através de todos os pontos e também na identificação de uma parcimoniosa estrutura de variância ao longo do tempo. As alternativas de modelagem podem ser aplicadas aos vários fatores aleatórios do modelo estatístico base. No contexto da estatística experimental com efeitos de tratamentos considerados fixos, estas modelagens, em geral se aplicam somente aos resíduos. Mas no caso do melhoramento, em que os tratamentos genéticos são considerados de efeitos aleatórios, essas modelagens podem ser aplicadas aos efeitos residuais e também genéticos. Inclusive, diferentes modelagens (modelo de repetibilidade, simetria composta - CS, auto-regressivo com variâncias heterogêneas - ARH, ante-dependência estruturado - SAD e multivariado) podem ser aplicadas aos efeitos genéticos e residuais, por exemplo, modelagem dos efeitos genéticos como ARH ou SAD e dos erros por um modelo multivariado. O presente trabalho compara formas alternativas de análise de medidas repetidas no melhoramento da produção de frutos de pinha. Vinte progênies de meias-irmãs foram avaliadas por três anos (2003, 2004 e 2005) no delineamento de blocos ao acaso com cinco repetições, onde cada parcela era constituída de quatro

plantas. A característica avaliada foi número de frutos por indivíduo. Todos os modelos foram analisados usando o software ASREML, com a estimação dos componentes de variância e a predição dos valores genéticos feita através do procedimento REML/BLUP. A comparação dos modelos ocorreu pelo teste de razão de verossimilhança e pelo critério de Akaike. A escolha do melhor modelo com base nas diferentes estruturas de covariâncias mostrou-se necessária na análise de medidas repetidas, visando maximizar a eficiência do melhoramento genético da pinha. O modelo SAD para os fatores progênie e parcela e multivariado para o fator resíduo mostrou-se a melhor abordagem para análise dos dados, propiciando eficiência e parcimônia em relação ao modelo multivariado completo. Com o modelo SAD foi possível a identificação de famílias superiores em cada colheita e também com maiores números totais de frutos.

## ABSTRACT

MARIGUELE, Keny Henrique, D.Sc., Universidade Federal de Viçosa, August, 2010.

**Methods of data analysis in longitudinal genetic improvement of custard apple (*Annona squamosa* L.).** Adviser: José Marcelo Soriano Viana. Co-advisers: Marcos Deon Vilela de Resende and Cosme Damião Cruz.

The pine cone or custard apple (*Annona squamosa* L.) is an important fruit crop in Brazil, especially in the Northeast. In this species, characters are repeatedly evaluated during the life of the organism and are called infinitely dimensional, in the sense that in each unit of time or age, the character can be measured, generating a set of multidimensional data. The interest in the analysis of such data usually resides in predicting values of individuals and progeny for particular point in time or through all points and in the identification of a parsimonious variance structure over time. The alternative of modeling can be applied to various random factors of the statistical model base. In the context of statistical experimental treatments considered fixed effects, these models, generally apply only to waste. But in case of improvement, in which treatments are considered genetic random effects, such modeling can be applied to the residual effects and also genetic. Also, different modeling (repeatability model, compound symmetry - CS, autoregressive with heterogeneous variance - ARH, ante-dependence structured - and multivariate SAD) can be applied to genetic effects and residuals, for example, modeling of genetic effects as ARH or SAD and mistakes by a multivariate model. This paper compares alternative ways of analyzing repeated measures to improve production of sugar apple. Twenty half-sibling progenies were evaluated over

three years (2003, 2004 and 2005) in a randomized block design with five replications. Each plot consisted of four plants. The trait was the number of fruits per individual. All models were analyzed using the software ASREML with the estimation of variance components and prediction of breeding values made by the REML / BLUP. The comparison of the models was the likelihood ratio test and the Akaike information criterion. Choosing the best model based on different covariance structures proved to be necessary in the analysis of repeated measures to maximize the efficiency of genetic improvement of cone. The SAD model for the factors and multiple offspring and parcel to the waste factor proved to be the best approach for data analysis, providing efficiency and parsimony over the full multivariate model. With the DSS model has been possible to identify superior families for each harvest and also with larger total numbers of fruit.

## 1. INTRODUÇÃO

A pinha ou fruta-do-conde (*Annona squamosa* L.) é uma fruteira importante para o Brasil, especialmente na Região Nordeste. É originária e foi domesticada no norte da América do Sul e América Central/Caribe (Ochse et al., 1974). Uma das limitações associadas ao cultivo da pinheira é a sua baixa variabilidade genética. Basicamente uma única variedade é cultivada em todas as regiões de produção (Manica et al., 2003; Pinto et al., 2005; São José, 1997).

Esses antecedentes apontam para a necessidade e importância de programas de melhoramento genético da espécie, especialmente aqueles baseados em introdução de novos indivíduos e progênies. Neste contexto, os testes de famílias de polinização aberta (meios irmãos) são fundamentais, pois propiciam ampla variabilidade genética tanto entre como dentro de famílias.

Na pinheira, os caracteres são avaliados repetidas vezes no decorrer da vida do organismo e são denominados infinitamente dimensionais, no sentido de que, em cada unidade de tempo ou idade, o caráter pode ser avaliado, gerando um conjunto multidimensional de dados. O interesse na análise desse tipo de dados geralmente reside na predição de valores dos indivíduos para determinado ponto no tempo ou através de todos os pontos e também na identificação de uma parcimoniosa estrutura de variância ao longo do tempo (Resende, 2007).

A análise de experimentos de medidas repetidas nas parcelas e/ou indivíduos apresenta peculiaridades pelo fato das várias colheitas serem correlacionadas entre si e pela possibilidade de haver heterogeneidade de variâncias e de covariâncias entre as várias colheitas ou safras. Um modelo completo e ótimo para analisar um conjunto de dados dessa natureza é o modelo multivariado (também denominado modelo com matriz de covariância não estruturada entre colheitas, UN) o qual trata cada colheita como se fosse uma variável diferente. E essa estrutura de covariância é aplicada a todos os fatores aleatórios do modelo tais quais os efeitos genotípicos de tratamento, efeitos

de parcelas e os efeitos residuais. Porém, considerando um número relativamente grande (três ou mais) de colheitas, tal modelo é difícil de ser ajustado (apresentando problema de convergência), além de ser superparametrizado, ou seja, depender da estimativa de um grande número de parâmetros (Resende, 2007).

No caso de medidas repetidas em cada indivíduo (ou tratamento) ao longo do tempo, várias alternativas (Resende e Sturion, 2001; Resende, 2002 e 2007, Resende e Thompson, 2003; Gilmour, 2006) existem para a predição de efeitos e modelagem da estrutura de correlação entre as referidas medidas. Geralmente as diferentes abordagens são implementadas via metodologia de modelos mistos.

De acordo com Resende et al. (2000), as equações de modelo misto permitem uma acurada predição de valores genéticos e genotípicos em situações de dados desbalanceados e com medidas repetidas, sendo prontamente aplicáveis e de grande utilidade prática em programas de melhoramento genético de espécies perenes.

O procedimento BLUP assume que os componentes de variância são conhecidos, sendo que as propriedades dos preditores somente são asseguradas nessas condições (Henderson, 1984). Assim, na prática, os componentes de variância devem ser estimados da maneira mais fidedigna possível, para que as estimativas possam substituir adequadamente os valores paramétricos (Resende et al., 2000). Nesse caso, o procedimento- padrão para estimação dos componentes de variância no contexto dos modelos lineares mistos é o da máxima verossimilhança restrita (REML), descrito em detalhes por Searle, Casella e Mc Culloch (1992) e Rao (1997).

No contexto dos modelos mistos, algumas opções de análise de medidas repetidas são: (i) modelo univariado simplificado de repetibilidade, o qual assume que o caráter é o mesmo (correlação genética igual a 1 através do tempo) de uma medição para outra, que as correlações fenotípicas (repetibilidades) são de iguais magnitudes entre todos os pares de idade e que as variâncias (genética e residual) são homogêneas; (ii) modelo univariado de repetibilidade mais interação genótipos x medições ou modelo de simetria composta (CS); (iii) modelo multivariado completo com matriz de covariância não estruturada, assumindo cada medida como um caráter diferente; (iv) modelo de regressão aleatória parcimonioso como aproximação do modelo multivariado; (v) ajuste de curva *spline* cúbica ou, alternativamente, *spline* tipo B, no intervalo de idades considerado; (vi) modelos processo caráter, tal como o modelo auto-regressivo com variâncias heterogêneas (ARH); (vii) modelos ante-dependência estruturados (SAD); (viii) modelo de correlação bandada (estrutura Toeplitz) com

correlações específicas para cada intervalo entre medições; (ix) estrutura de simetria composta com variâncias heterogêneas (CSH).

O uso do modelo de repetibilidade (o mais simples e parcimonioso) em lugar do modelo multivariado (o mais completo e complexo) conduz a grande eficiência, obtendo-se perdas de apenas 0 a 5%, quando a correlação genética entre medições sucessivas é alta (acima de 0,80). Nesses casos, a correlação entre ordenamentos pelo modelo de repetibilidade e pelo modelo multivariado é muito próxima de 1 (Mrode, 2005). O grande benefício do modelo de repetibilidade comparado com o multivariado refere-se à menor demanda computacional e à necessidade de estimativas de poucos parâmetros genéticos. Um modelo misto de repetibilidade, o qual contempla simultaneamente a herdabilidade e a repetibilidade do caráter foi proposto por Resende et al.(2000) para o caso do delineamento experimental de blocos ao acaso com várias plantas por parcela. No entanto, estruturas intermediárias entre os modelos de repetibilidade e multivariado podem ser mais eficientes.

No melhoramento animal, os dados multidimensionais associados a lactações, crescimento em peso, produção de lã e produção de ovos, geralmente são analisados via regressão aleatória, *splines*, modelos processo caráter ou SAD. Gilmour (2006) destacam os dois últimos como mais eficientes e relata que as propriedades matemáticas da regressão polinomial revelam que o ajuste de *splines* cúbicas é mais eficiente que o uso da técnica de regressão aleatória.

Em plantas perenes, o número de medições realizadas varia tipicamente de 3 a 6 safras anuais, pois um número maior de safras anuais compromete a eficiência dos programas de melhoramento por unidade de tempo. Com número de medições desta ordem, as técnicas de regressão aleatória e de *splines* tendem a não ser eficientes devido ao reduzido número de idades abrangido pelos dados. Tais técnicas são muito empregadas no melhoramento animal, em que indivíduos de diferentes idades (as medidas repetidas não ocorrem em intervalos fixos) são avaliados produzindo grande número de pontos em termos de idades. Assim, as opções mais interessantes aos melhoristas de plantas perenes são os modelos de repetibilidade (quando as suposições são atendidas), repetibilidade + interação genótipos x medições (CS), ARH, SAD e CSH.

A estrutura de covariância de simetria composta (CS) assume tanto homogeneidade de variâncias quanto de covariâncias entre colheitas. Essa abordagem é desejável porque depende do menor número de possível de parâmetros. No entanto,

pode ser ineficiente no caso de grande heterogeneidade de variância e de covariância entre colheitas, algumas vezes decorrentes de efeitos de escala de observação de uma colheita a outra. Para uma aplicação segura dessa estrutura CS, recomenda-se, em caso de presença de grande heterogeneidade de variâncias, a correção prévia dos dados de cada colheita por meio da multiplicação dos mesmos pela razão  $h_i/h_m$ , em que  $h_i$  refere-se à raiz quadrada da herdabilidade na colheita  $i$  e  $h_m$  refere-se à raiz quadrada da média das herdabilidades nas várias colheitas (Resende, 2007). Esse procedimento considera simultaneamente a heterogeneidade de variância genética e fenotípica entre colheitas, penalizando pela alta variância fenotípica e capitalizando pela alta variância genética, o que equivale aproximadamente a padronizar pelo desvio padrão ambiental de cada colheita e multiplicar pela média desses desvios padrões. Procedendo-se dessa forma, a aplicação de estruturas CS fornece resultados semelhantes aqueles que se obtêm quando se aplica a estrutura de simetria composta com variâncias heterogêneas (CSH).

Essas alternativas (ARH, SAD e multivariado) de modelagem podem ser aplicadas aos vários fatores aleatórios do modelo estatístico base. No contexto de estatística experimental com efeitos de tratamentos considerados fixos, estas modelagens, em geral se aplicam somente aos resíduos. Mas no caso do melhoramento, em que os tratamentos genéticos são considerados de efeitos aleatórios, essas modelagens podem ser aplicadas aos efeitos residuais e também genéticos. Inclusive, diferentes modelagens podem ser aplicadas aos efeitos genéticos e residuais. Para algumas espécies, a modelagem dos efeitos genéticos como ARH ou SAD e dos erros por um modelo multivariado, mostram-se eficientes (Resende et al., 2003).

O objetivo desse trabalho foi comparar formas alternativas de análise de medidas repetidas no melhoramento genético de pinha. Diferentes estruturas de matriz de covariância para os fatores progênie, parcela e resíduo foram avaliadas em um teste de progênies de meias-irmãs de pinheiras (*Annona squamosa* L.) avaliadas em três anos. O trabalho visa fornecer subsídios para a realização de eficientes programas de melhoramento da espécie e contribuir para melhorar o sistema produtivo dessa cultura.

## 2. MATERIAL E MÉTODOS

### Obtenção das progênies

As progênies foram obtidas em três locais: Aracati-CE, Mossoró-RN e Serra do Mel-RN. Em cada um destes locais, foram visitados alguns pomares caseiros e, com base na sanidade e vigor das plantas, rendimento aparente de frutos e tipo de frutos (boa conformação e bom tamanho), foi selecionado as matrizes. Um total de 25 matrizes foi selecionado. De cada matriz, foram colhidos seis frutos que, após a maturação sob condições ambientais, foram abertos para a coleta das sementes. As sementes foram lavadas e postas para secar a sombra, escolhendo-se 30 sementes de cada matriz.

As plantas foram conduzidas em haste única até 0,40 m. Ficando com três ramos para a formação da copa. As irrigações foram efetuadas por microaspersão três vezes por semana durante 2 horas. Foram realizadas podas de limpeza a cada três meses, capina e roço a cada quatro meses.

### Delineamento experimental

As vinte progênies foram avaliadas por três anos (2003, 2004 e 2005) no delineamento de blocos ao acaso com cinco repetições, onde cada parcela era constituída de quatro plantas. A característica avaliada foi número de frutos por indivíduo.

### Modelos

Em todos os modelos de análise, os fatores ano, repetição e a interação repetição x ano foram considerados como efeitos fixos. Os demais efeitos foram tratados como aleatórios.

#### a) Modelo de Repetibilidade

$y = Xb + Zg + Wc + Tp + e$ , em que  $y$  é o vetor de dados,  $b$  é o vetor dos efeitos das combinações ano-repetição (assumidos como fixos) somados à média geral,  $g$  é o vetor dos efeitos genotípicos de progênies (assumidos como aleatórios),  $c$  é vetor dos efeitos

de ambiente permanente (parcelas no caso) (aleatórios),  $p$  é vetor dos efeitos permanentes de indivíduo (aleatórios) e  $e$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

### Distribuições e estruturas de médias e variâncias

$$y|b,V \sim N(Xb, V)$$

$$g| \sigma_g^2 \sim N(0, \sigma_g^2)$$

$$c| \sigma_c^2 \sim N(0, I \sigma_c^2)$$

$$p| \sigma_p^2 \sim N(0, I \sigma_p^2)$$

$$e| \sigma_e^2 \sim N(0, I \sigma_e^2)$$

$$Cov(g, c') = 0; \quad Cov(g, p') = 0; \quad Cov(g, e') = 0;$$

$$Cov(p, c') = 0; \quad Cov(p, e') = 0; \quad Cov(c, e') = 0$$

ou seja:

$$E \begin{bmatrix} y \\ g \\ c \\ p \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} y \\ g \\ c \\ p \\ e \end{bmatrix} = \begin{bmatrix} V & ZG & WC & TP & R \\ GZ' & G & 0 & 0 & 0 \\ CW' & 0 & C & 0 & 0 \\ PT' & 0 & 0 & P & 0 \\ R & 0 & 0 & 0 & R \end{bmatrix}, \text{ em que:}$$

$$G = I \sigma_g^2$$

$$P = I \sigma_p^2$$

$$C = I \sigma_c^2$$

$$R = I \sigma_e^2$$

$$V = Z\sigma_g^2 Z' + W I \sigma_c^2 W' + T I \sigma_p^2 T' + I \sigma_e^2.$$

A estrutura da matriz de covariância entre medidas repetidas para o fator progênie é da

$$\text{forma } \Sigma = \begin{bmatrix} \sigma_g^2 & \sigma_g^2 & \sigma_g^2 \\ & \sigma_g^2 & \sigma_g^2 \\ & & \sigma_g^2 \end{bmatrix} \text{ e a estrutura da matriz de correlação para o fator progênie é}$$

$$\text{da forma } H = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}.$$

### b) Modelo de Simetria Composta – CS (repetibilidade mais interação gxa)

$y = Xb + Zg + Wp + Tp + Qi + e$ , em que  $y$  é o vetor de dados,  $b$  é o vetor dos efeitos das combinações ano-repetição (assumidos como fixos) somados à média geral,  $g$  é o vetor dos efeitos genotípicos de progênies (assumidos como aleatórios),  $c$  é vetor dos efeitos de ambiente permanente (parcelas no caso) (aleatórios),  $p$  é vetor dos efeitos

permanentes de indivíduo (aleatórios),  $i$  é o vetor dos efeitos da interação genótipos x anos e  $e$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

A estrutura de covariância é similar à anterior, porém com o acréscimo da matriz de covariância dos efeitos da interação,  $GA = I \sigma_{ga}^2$ . A estrutura da matriz de covariância

para o fator progênie é da forma  $\Sigma = \begin{bmatrix} \sigma_g^2 + \sigma_{gxa}^2 & \sigma_g^2 & \sigma_g^2 \\ & \sigma_g^2 + \sigma_{gxa}^2 & \sigma_g^2 \\ & & \sigma_g^2 + \sigma_{gxa}^2 \end{bmatrix}$  e a estrutura da

matriz de correlação entre medidas repetidas para o fator progênie é da forma

$$H = \begin{bmatrix} 1 & \sigma_g^2 / (\sigma_g^2 + \sigma_{gxa}^2) & \sigma_g^2 / (\sigma_g^2 + \sigma_{gxa}^2) \\ & 1 & \sigma_g^2 / (\sigma_g^2 + \sigma_{gxa}^2) \\ & & 1 \end{bmatrix}.$$

As estruturas ARH, SAD e Multivariada (UN) são descritas a seguir com base em Resende (2007). Em todas essas abordagens foram contemplados os efeitos de progênie, de parcela e de resíduos.

### **c) Modelo auto-regressivo com variâncias heterogêneas (estrutura auto-regressiva) – ARH**

Pletcher e Geyer (1999) sugeriram o uso dos modelos processo caráter para análise de medidas repetidas. Esses modelos são baseados na teoria de processos estocásticos e foram estendidos por Jaffrezic e Pletcher (2000) visando relaxar a suposição mais restritiva de estacionariedade das correlações. O mais simples modelo processo caráter usa a função de covariância  $C(t,s) = \sigma_t \sigma_s \rho^{(t-s)}$ , em que  $C(t,s)$  é a covariância entre medições nas safras ou tempos  $t$  e  $s$ ,  $\sigma_t$  é o desvio padrão do caráter na safra  $t$  e  $\rho^{(t-s)}$  é a correlação entre medições nas safras ou tempos  $t$  e  $s$ . Quando os dados são coletados em intervalos igualmente espaçados, esse processo caráter equivale ao modelo autorregressivo com variâncias heterogêneas (ARH). Assim, para medidas repetidas em plantas perenes, em que geralmente têm-se avaliações em intervalos regulares, os modelos de processo caráter e ARH são equivalentes.

A estrutura da matriz de covariância é:

$$\Sigma = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_1 \sigma_2 \rho^{(1-2)} & \sigma_1 \sigma_3 \rho^{(1-3)} \\ & \sigma_{g_2}^2 & \sigma_2 \sigma_3 \rho^{(2-3)} \\ & & \sigma_{g_3}^2 \end{bmatrix}$$

Este modelo assume correlações diferentes entre idades e reconhece a existência de correlações serial entre medidas repetidas. A estrutura geral de correlação envolvendo três diferentes safras é:

$$H = \begin{bmatrix} 1 & \rho^{[t_2-t_1]} & \rho^{[t_3-t_1]} \\ & 1 & \rho^{[t_3-t_2]} \\ & & 1 \end{bmatrix}$$

Verifica-se que o modelo AR estima uma só correlação  $\rho$  e projeta-a para os demais lags. Neste caso, a matriz de covariância genética é dada por  $G=SRS$  em que  $S$  é uma matriz diagonal com elementos equivalentes à raiz quadrada da variância genética em cada colheita.

#### **d) Modelo ante dependência estruturado (SAD)**

A idéia básica dos modelos ante dependência (AD) é que uma observação no tempo  $t$  pode ser explicada por observações prévias. Os valores da observação em determinado tempo dependem da observação no tempo imediatamente anterior e mais uma nova variação, ou seja, mais um acréscimo ocorrido no intervalo entre as duas medições. Nunez-Anton e Zimmermann (2000) propuseram modelo ante dependência estruturado (SAD) no qual o número de parâmetros é menor do que nos tradicionais modelos AD. Esses modelos podem lidar com padrões de correlação altamente não estacionários e correspondem, em suas especificações mais simples, a uma generalização dos modelos autorregressivos. Consideram também a heterogeneidade de variância entre medições. O SAD é também favorável em termos do número de parâmetros ajustados o qual é geralmente pouco maior do que o número ajustado pelos modelos ARH. A matriz de covariância é da forma:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho_2 \\ & & \sigma_3^2 \end{bmatrix}$$

A matriz de correlação é da forma:

$$H = \begin{bmatrix} 1 & \rho_1 & \rho_1\rho_2 \\ & 1 & \rho_2 \\ & & 1 \end{bmatrix}$$

Verifica-se que a correlação (ou distância) entre as medidas 1 e 3 equivale ao produto entre a correlação da medida 1 com a 2 ( $\rho_1$ ) e a correlação da medida 2 com a 3 ( $\rho_2$ ).

#### e) Modelo multivariado (matriz de covariância não estruturada UN)

Este é o modelo mais completo, utiliza toda a informação simultaneamente e trata safras como sendo caracteres diferentes e correlacionados, considerando suas diferentes herdabilidades e correlações genéticas.

A estrutura geral da matriz de covariância é:

$$\Sigma = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{12} & \sigma_{13} \\ & \sigma_{g_2}^2 & \sigma_{23} \\ & & \sigma_{g_3}^2 \end{bmatrix}$$

A estrutura geral de correlação entre três diferentes safras é dada por:

$$H = \begin{bmatrix} 1 & a & b \\ & 1 & c \\ & & 1 \end{bmatrix}, \text{ em que } a, b \text{ e } c \text{ representam diferentes valores de correlação.}$$

Neste caso, a matriz H é denominada não estruturada, e o modelo é superparametrizado e proibitivo na prática quando muitas safras são consideradas. Sob modelo multivariado surgem problemas de estimação (matrizes não positivas definidas), principalmente quando o valor paramétrico encontra-se próximo ao limite do espaço do parâmetro.

#### Comparação dos modelos

Todos os modelos foram analisados usando o software ASREML (Gilmour e Thompson, 1998; Gilmour et al., 2002), com a estimação dos componentes de variância feita através de REML/BLUP.

Na análise de modelos mistos com dados desbalanceados, os efeitos do modelo não são testados via teste F tal como se faz no modelo da análise de variância. Nesse caso, para efeitos aleatórios, o teste cientificamente recomendado é o teste da razão de verossimilhança (LRT) (Resende, 2007).

### **a) Teste de Razão de Verossimilhança (LRT)**

Foi usada a inferência probabilística ou critério de Deviance (DIC). Além de sua utilidade na estimação, o princípio de verossimilhança também permite comparar a adequabilidade de vários modelos, desde que tenham uma estrutura hierárquica ou aninhada (Resende, 2007).

Os seguintes passos foram adotados:

- Obtenção do ponto de máximo do logaritmo da função de verossimilhança residual (Log L) para modelos a serem comparados;
- Obtenção da deviance  $D = -2 \text{ Log L}$  para os dois modelos.
- Obtenção da diferença entre as deviances para ter-se o qui-quadrado calculado e realizar o teste da razão de verossimilhança (LRT);
- Teste da significância dessa diferença usando o teste qui-quadrado com o número de graus de liberdade igual à diferença entre o número de parâmetros entre os dois modelos testados.

### **b) Critério de Informação de Akaike (AIC)**

Esse critério penaliza a verossimilhança pelo número de parâmetros independentes ajustados. Por esse critério, qualquer parâmetro extra, deve aumentar a verossimilhança por ao menos uma unidade para que o mesmo entre no modelo. O AIC é dado por  $AIC = -2 \text{ Log L} + 2p$ , em que  $p$  é o número de parâmetros estimados (Akaike, 1974), conforme Littell et al. (1998).

### 3. RESULTADOS E DISCUSSÃO

A comparação entre os modelos de análise visou encontrar um modelo que melhor represente os dados e seja parcimonioso. Isto foi feito empregando-se o teste da razão de verossimilhança e o critério de Akaike (AIC). O teste de razão de verossimilhança compara dois modelos, onde um dos modelos é uma versão restrita do outro, ou seja, um dos modelos tem 'n' parâmetros adicionais. O teste verifica se esses parâmetros adicionais melhoram significativamente o modelo. O AIC é baseado na teoria de decisão e penaliza os modelos com número grande de parâmetros para evitar excesso de parametrização. Para a tomada de decisão sobre qual modelo selecionar é necessário que o AIC seja calculado para todos os modelos considerados e escolhido aquele com menor valor.

Os resultados das diferentes modelagens são apresentados nas Tabelas de 1 a 4. A comparação entre os modelos de repetibilidade e CS mostrou que o segundo apresentou a menor deviance (6346,42) (Tabela 1). A escolha desse modelo foi confirmada pelo AIC, que apresentou os valores de 6403,86 e 6356,42, respectivamente. Como esperado, a estimativa da  $\sigma_g^2$  no modelo mais simples foi superior (5,77) a obtida no modelo CS (2,98); pois no modelo simples de repetibilidade ocorre à superestimação da mesma devido à interação estar embutida. O teste de

$\chi^2 = 49,44^{**}$  foi significativo a 1% de probabilidade, mostrando diferença entre os dois modelos é compensador acrescentar um parâmetro a mais ( $\sigma_{g,xi}^2$ ).

Comparando o modelo CS com o ARH, este forneceu uma deviance menor (6306,44) (Tabela 2). O índice de AIC foi de 6334,44 também inferior ao obtido no outro modelo (6356,42). O teste de  $\chi^2 = 39,98^{**}$  mostrou diferença significativa entre os modelos ao nível de significância de 1% de probabilidade, sendo as estimativas obtidas por este modelo mais confiáveis, apesar do aumento do número de parâmetros.

Os anos três, um e dois apresentaram variabilidade genética decrescente. No caso do ARH (Tabela 2), as baixas estimativas de variâncias das parcelas conduziram a correlações entre os anos iguais a 1, não sendo assim confiáveis. As maiores estimativas de correlação genética foram obtidas entre as colheitas 1 e 2 (0,57; 0,78 e 0,90) pelos modelos ARH, SAD e multivariado, respectivamente (Tabelas 2, 3 e 4).

A deviance obtida pelo SAD foi de 6137,36 (Tabela 3) e inferior a obtida pelo ARH. Mais uma vez o AIC de 6169,36 foi inferior ao obtido pelo SAD (6334,44). O teste  $\chi^2 = 169,08^{**}$  revelou diferença significativa entre os modelos, e que apesar do aumento de dois parâmetros as estimativas obtidas por este modelo são mais confiáveis.

Os modelos SAD e multivariado apresentaram quase a mesma deviance (6137,36 e 6136,76) (Tabelas 3 e 4, respectivamente). Dever-se-ia escolher o modelo multivariado por apresentar valor ligeiramente menor da deviance. Entretanto, os valores de AIC foram 6169,36 e 6172,72 e, portanto, menor para o SAD. No entanto, seguindo o critério da parcimônia, confirmado pelo teste  $\chi^2 = 0,6^{ns}$ , não existe diferença significativa entre os modelos. Sendo assim, prefere-se o modelo menos parametrizado, que no caso é o SAD que apresenta dois parâmetros a menos.

Adicionalmente, o modelo multivariado completo (não estruturado) tende a apresentar problemas de convergência em virtude das altas correlações geralmente

verificadas entre medidas repetidas e do elevado número de parâmetros a ser estimado. A abordagem multivariada torna-se proibitiva quando o número de colheitas ou safra é elevado. Algumas vezes, mesmo com pequeno número de safras, as matrizes de covariância não estruturadas não são positivas definidas e conduzem a correlações maiores que 1. Medidas repetidas altamente correlacionadas aumentam o risco de obtenção de matrizes de covariância não positivas definidas e de não convergência no processo de estimação de componentes de variância. Também, quanto maior a ordem da matriz de covariância, maior é a chance de que ela não seja positiva definida (Resende, 2007).

Com correlações genéticas muito discrepantes entre pares de colheitas e grande número de colheitas, outras estruturas intermediárias entre UN e CS podem ser usadas. Para caracteres de plantas perenes, em que as correlações seguem alguns padrões em função das distâncias entre as medidas consideradas, em geral havendo decréscimo das correlações em função do aumento dessas distâncias, os modelos ARH e SAD tendem a ser os mais adequados (Resende e Thompson, 2003; Resende et al., 2006). O modelo ARH pode ser também ser adequado nas situações em que as correlações apresentam a mesma magnitude em intervalos de mesma dimensão, por exemplo, correlações iguais entre as colheitas 1 e 2 e entre as colheitas 2 e 3. Em geral, com a estabilização do caráter, a correlação entre as colheitas 2 e 3 tende a ser maior que a correlação entre as colheitas 1 e 2, e portanto, o modelo SAD parece ser mais coerente pois contempla isso.

Cecon et al. (2008) trabalhando com a produção de cinco anos de 50 clones de café 'Conilon' verificaram que o modelo que proporcionou o melhor ajuste foi o CSH. Freitas (2008) estudou 28 genótipos de cana de açúcar por três anos e verificou que o modelo de matriz de covariância apropriado foi o multivariado não estruturado, escolhido pelo critério de AIC. Enquanto que Resende et al. (2006) trabalharam com 1800 indivíduos de erva-mate em três colheitas e concluíram que os modelos ARH,

SAD e multivariado apresentaram basicamente a mesma deviance mas o ARH foi o escolhido por ser mais parcimonioso pelo AIC.

Na literatura há escassez de informação sobre a estimação de parâmetros genéticos e dados de medidas repetidas em fruteiras, embora a importância dos modelos de repetibilidade para avaliação genética em algumas espécies perenes tenha sido enfatizada (Resende et al., 2000). Assim, os resultados obtidos nesse trabalho contribuem para preencher uma lacuna e propiciam informações sobre modelagens adicionais que podem ser aplicadas às fruteiras. As predições dos valores genéticos das famílias, obtidas pelo modelo escolhido (SAD) são apresentadas na Tabela 5.

O teste de famílias é uma das estratégias mais exploradas para a seleção em espécies perenes (Oliveira et al., 2006). Praticando essa seleção, considerando uma intensidade de seleção de 25% das famílias, observa-se que na colheita 1 as melhores famílias foram a 3, 4, 15, 11 e 10, na colheita 2 (3, 4, 11, 10 e 2) e na colheita 3 (11, 6, 20, 8 e 4). Considerando as médias das colheitas, as melhores famílias foram a 11, 4, 6, 3 e 15. Sendo que as famílias 4 e 11 estiveram entre as que apresentaram os maiores valores genéticos nas três colheitas individuais (Tabela 5).

#### **4. CONCLUSÃO**

- a) A escolha do melhor modelo com base nas diferentes estruturas de covariâncias mostrou-se necessária na análise de medidas repetidas, visando maximizar a eficiência do melhoramento genético da pinha.
- b) O modelo SAD para os fatores progênie e parcela e multivariado para o fator resíduo mostrou-se a melhor abordagem para análise dos dados, propiciando eficiência e parcimônia em relação ao modelo multivariado completo.
- c) Com o modelo SAD foi possível a identificação de famílias superiores em cada colheita e também com maiores produtividades totais de frutos.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- Akaike, H. A new look at the statistical model identification. **IEEE transactions on Automatic Control**, v.19, p.716-723, 1974.
- Cecon, P.R; Fabyano Silva, F.F. e; Ferreira, A.; Ferrão, R.G., Carneiro, A.P.S.; Detmann, E.; Faria, P.N.; Morais, e T.S. da S. Análise de medidas repetidas na avaliação de clones de café 'Conilon'. **Pesquisa Agropecuária Brasileira**, Brasília, v.43, n.9, p.1171-1176, set. 2008.
- Freitas, E. G. de Análise de dados longitudinais em experimentos com cana-de-açúcar. Piracicaba, 2007, 75 p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- Gilmour, A. R. Statistical models for multidimensional (longitudinal/spatial) data. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006. **Proceedings**. Belo Horizonte, Intituto Prociência, 2006 CD-ROM.
- Gilmour, A. R.; Thompson, R.; Cullis, B. R.; Welham, S. J. ASREML estimates variance matrices from multivariate data using the animal model. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. **Proceedings**. Paris: INRA, 2002. I CD-ROM (Communication, n. 28).

- Gilmour, A. R.; Thompson, R. Modelling variance parameters in ASREML for repeated measures. In: WORLD CONGRESS ON GENETIC APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. **Proceedings**. Armidale: AGBU: University of New England, 1998, v. 27, p. 453-454.
- Henderson, C.R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462p.
- Jaffrezic, F.; Pletcher, S. D. Statistical models for estimating the genetic basis of repeated measures and other function valued traits. **Genetics**, v. 156, p. 913-922, 2000.
- Littell, R.C.; Henry, P.R.; Ammerman, C.B. Statistical analysis of repeated measures using SAS procedures. **Journal of Animal Science**, v.76, p.1216-1231, 1998.
- Manica, I.; Icuma, I.M.; Junqueira, K.P.; Oliveira, M.A.S.; Cunha, M.M.; Oliveira JR., M.E.; Junqueira, N.T.V; Alves, R.T. Frutas Anonáceas (ata ou pinha, atemólia, cherimólia e graviola). **Tecnologia de produção, pós-colheita, mercado. Cinco Continentes Editora**. 596p. 2003.
- Mrode, R. A. **Linear models for the prediction of animal breeding values**. 2<sup>nd</sup> ed. Wallingforf: CAB Intermtional, 2005, 386 p.
- Nunez-Anton, V.; Zimmermann, D. L. Modelling non-stationary longitudinal data. **Biometrics**, v. 56, p. 699-705, 2000.
- Oliveira, A. N.; Silva, A. C.; Rosado, S. C. S.; Rodrigues, E. A. C. Variações genéticas para características do sistema radicular de mudas de baru (*Dipteryx alata* Vog.). **Revista Árvore**, Viçosa, v. 30, n. 6, p. 905-909, 2006.
- Ochse, J.J.; Soule JR., M.J.; Dijkman, M.J.; Wehlburg, C. Otros cultivos frutales. In: **Cultivo y mejoramiento de plantas tropicales e subtropicales**. México: Ed. Limusa, 1974. p.587-818.

Pinto, A.C.Q.; Cordeiro, M.C.R.; Andrade, S.R.M.; Ferreira, F.R.; Filgueiras, H.A.C.; Alves, R.E.; Kimpara, D.J. **Fruits for the future 5: Annona species**. International Centre for Underutilised Crops, University of Southampton, Southampton: IPGRI, 2005. 263p.

Pletcher, S. D.; Geyer, C. J. The genetic analysis of age-dependent traits: modelling a character process. **Genetic**, v. 153, p. 825-833, 1999.

Rao, P.S. **Variance components: mixed models methodologies and applications**. Boca Raton: CRC Press, 1997. 350p.

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

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

Resende, M. D. V. de; Simeão, R. M.; Fernandes, J. S. C. Blup individual com medidas repetidas aplicado ao melhoramento de espécies perenes. **Ciência e Agrotecnologia**, Lavras, v.24, n.4, p.986-997, 2000.

Resende, M. D. V. de; Sturion, J. A. **Análise genética de dados com dependência espacial e temporal no melhoramento de plantas perenes via modelos geoestatístico e de series temporais empregando REML/BLUP ao nível individual**. Colombo. Embrapa Florestas, 2001, 80 p.

Resende, M. D. V. de; Thompson, R. **Multivariate spatial statistical analysis of multiple experiments and longitudinal data**. Colombo: Embrapa Florestas, 2003, 126 p. (Embrapa Florestas. Documento, 90).

Resende, M. D. V. de; Thompson, R.; Welham, S. J. Multivariate spatial statistical analysis of longitudinal data in perennial crops. **Brazilian Journal of Mathematics and Statistics**, v. 24, n. 1, p. 147-169, 2006.

Resende, M. D. V. de; Thompson, R.; Welham, S. J. Multivariate spatial statistical analysis of longitudinal data in perennial crops. In: INTERNATIONAL BIOMETRIC SOCIETY CONFERENCE – BRITISH REGION, Reading. **Proceedings**. Reading: University of Reading, School of Applied Statistics, 2003, p. 70-71.

São José, A.R. Aspectos generales de las anonáceas em Brasil In: **CONGRESO INTERNACIONAL DE ANONÁCEAS**, 1., 1997, Chapingo. Memorias... Chapingo: Universidad Autónoma Chapingo, p. 92-103, 1997.

Searle, S.R.; Casella, G.; McCulloch, C.E. **Variance components**. New York: J. Wiley, 1992. 528p.

Tabela 1 – Estimativas das variâncias associadas ao modelo de repetibilidade e simetria composta (CS), aplicado aos dados originais de número de frutos em três colheitas de *Annona squamosa* L.. Número total de parâmetros de cada modelo iguais a 4 e 5 respectivamente.

Variância	Modelos	
	Repetibilidade	CS
$\sigma_g^2$	5,77	2,98
$\sigma_p^2$	2,97	2,96
$\sigma_i^2$	7,52	10,21
$\sigma_{gxa}^2$	-	8,38
$\sigma^2$	67,45	59,38
<b>Deviance</b>	<b>6395,86</b>	<b>6346,42</b>

Tabela 2 – Estimativas dos parâmetros de variâncias, covariâncias e correlações associados ao modelo processo caráter, denominado autorregressivo de primeira ordem com variâncias heterogêneas (ARH), aplicado aos dados originais de número de frutos em três colheitas de *Annona squamosa* L.. Numero total de parâmetros igual a 14.

<b>Tratamento (progênie)</b>			<b>Parcela</b>			<b>Resíduo</b>		
<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>	<b>Covar\Variância\Corre</b>
8,85	0,57	0,32	0,99E-13	-	-	67,18	-0,06	-0,01
2,58	2,31	0,57	0,32E-13	0,10E-13	-	-4,14	74,26	0,09
4,46	4,06	21,94	0,48E-6	0,15E-6	22,56	-0,66	5,86	51,59
<b>Deviance = 6306,44</b>								

Tabela 3 – Estimativas dos parâmetros de variâncias, covariâncias e correlações associados ao modelo ante dependência estruturado (SAD), aplicado aos dados originais de número de frutos em três colheitas de *Annona squamosa* L.. Número total de parâmetros igual a 16.

Tratamento (progênie)			Parcela			Resíduo		
Covar	Variância	Corre	Covar	Variância	Corre	Covar	Variância	Corre
8,51	0,78	0,24	5,98	-0,41	-0,20	47,64	0,26	0,16
4,04	3,18	0,31	-1,71	3,00	0,49	10,28	33,56	0,24
3,18	2,50	20,81	-1,76	3,09	12,98	12,09	15,09	116,83
<b>Deviance = 6137,36</b>								

Tabela 4 – Estimativas dos parâmetros de variâncias, covariâncias e correlações associados ao modelo multivariado completo aplicado aos dados originais de número de frutos em três colheitas de *Annona squamosa* L.. Número total de parâmetros igual a 18.

Tratamento (progênie)			Parcela			Resíduo		
Covar	Variância	Corre	Covar	Variância	Corre	Covar	Variância	Corre
6.31	0.90	0.35	1.01	-0.76	-0.27	47.62	0.26	0.15
3.35	2.19	0.22	-0.52	0.46	0.75	10.42	33.72	0.24
3.32	1.25	14.3	-0.38	0.72	1.99	11.34	14.74	116.3
<b>Deviance = 6136,76</b>								

Tabela 5 – Valores genéticos obtidos a partir das estimativas do modelo SAD para o número de frutos das vinte progênies de meias-irmãs de pinheiras (*Annona squamosa* L.).

Famílias	Colheita 1	Colheita 2	Colheita 3	Geral
1	10,9583	12,0233	7,5383	10,1733
2	14,8385	15,4516	14,3663	14,8855
3	20,5083	17,9033	11,6853	16,6990
4	19,8643	16,8733	17,3113	18,0163
5	15,1084	14,5757	15,0277	14,9039
6	14,8927	15,4056	20,1953	16,8312
7	12,1393	13,0583	10,9493	12,0490
8	15,1880	13,4803	17,8943	15,5209
9	11,4383	12,6753	15,8423	13,3186
10	16,0593	15,8003	13,3953	15,0850
11	16,2263	16,5863	23,9203	18,9110
12	15,3689	14,9183	14,3662	14,8845
13	14,1159	14,9596	13,2643	14,1133
14	13,9403	14,7173	14,2993	14,3190
15	16,4583	14,9395	15,6724	15,6901
16	15,4703	15,2915	14,2862	15,0160
17	11,3553	12,9923	15,4082	13,2519
18	12,2683	14,2471	7,6353	11,3836
19	14,1521	14,3123	12,5523	13,6722
20	13,8149	13,9540	18,5553	15,4414

## **ANEXOS**

**Anexo 1** – Programa ASREML para o modelo de repetibilidade

PG1

Bloco 5

Progenie 20

Parcela 100

Idade 4

Indiv 399

NFrutos

Pinha3Rep.txt !SKIP 1 !REPEAT !MAXIT 100

NFrutos ~ Idade Idade.Bloco !r Parcela Progenie Indiv

**Anexo 2** – Programa ASREML para o modelo de simetria composta

PG1

Bloco 5

Progenie 20

Parcela 100

Idade 4

Indiv 399

Interação GM 60

NFrutos

Pinha3RepGM.txt !SKIP 1 !REPEAT !MAXIT 100

NFrutos ~ Idade Idade.Bloco !r Parcela Progenie Indiv GM

**Anexo 3** – Programa ASREML para o modelo autorregressivo de primeira ordem com variâncias heterogêneas (ARH)

PG1

Bloco 5

Progenie 20

Parcela 100

Idade 3

NFrutos

Pinha3.txt !SKIP 1 !REPEAT !MAXIT 100

NFrutos ~ Idade Idade.Bloco !r Idade.Parcela Idade.Progenie

1 2 2 !ASMV 3

399

Idade 0 US !+6

42.38

-6.343 42.29

-8.204 -5.49 47.20

Idade.Parcela 2

Idade 0 ARH !+3

0.3156

6.258 5.782 38.81

Parcela

Idade.Progenie 2

Idade 0 ARH !+3 !GP

0.6346

8.25 3.75 21.23

Progenie

**Anexo 4** – Programa ASREML para o modelo ante dependência estruturado (SAD)  
PG2  
Bloco 5  
Progenie 20  
Parcela 100  
Planta 400  
NFrutos2  
NFrutos3  
NFrutos4  
PinhaNF3.txt !SKIP 1 !MAXIT 400  
NFrutos2 NFrutos3 NFrutos4 ~ Trait.mu Trait.Bloco !r Trait.Progenie Trait.Parcela  
1 2 2  
0  
Trait 0 US 4.22 .1 11.88 1.0 2 24.15  
Trait.Progenie 2  
Trait 0 ANTE 8.4 .05 5.56 .11 .3 1.26  
Progenie  
Trait.Parcela 2  
Trait 0 ANTE 8.4 .01 5.56 .02 .11 .15  
Parcela

**Anexo 5** – Programa ASREML para o modelo multivariado completo

PG2

Bloco 5

Progenie 20

Parcela 100

Planta 400

NFrutos2

NFrutos3

NFrutos4

PinhaNF3.txt !SKIP 1 !MAXIT 400

NFrutos2 NFrutos3 NFrutos4 ~ Trait.mu Trait.Bloco !r Trait.Progenie Trait.Parcela

1 2 2

0

Trait 0 US 4.22 .1 11.88 1.0 2 24.15

Trait.Progenie 2

Trait 0 US 8.4 .05 5.56 .11 .3 1.26

Progenie

Trait.Parcela 2

Trait 0 US 8.4 .01 5.56 .02 .11 .15

Parcela