



DOCUMENTOS, 47

ISSN 1517-536X

**ANÁLISE ESTATÍSTICA DE MODELOS MISTOS  
VIA REML/BLUP NA EXPERIMENTAÇÃO EM  
MELHORAMENTO DE PLANTAS PERENES**

Marcos Deon Vilela de Resende

Colombo  
2000



Exemplares desta publicação podem ser solicitados à:

***Embrapa Florestas***

Estrada da Ribeira km 111 - Caixa Postal 319

83411-000 - Colombo, PR Brasil

Fone: (0\*\*41) 666-1313

Fax: (0\*\*41) 666-1276

E-mail: sac@cnpf.embrapa.br

Tiragem: 300 exemplares

**Comitê de Publicações:**

Américo Pereira de Carvalho, Antônio Carlos de S. Medeiros, Edilson Batista de Oliveira, Erich Gomes Schaitza, Guiomar Moreira de Souza Braguiná (Secretaria Executiva), Honorino Roque Rodigheri, Jarbas Yukio Shimizu, José Alfredo Sturion, Moacir José Sales Medrado (Presidente), Patrícia Póvoa de Mattos, Rivail Salvador Lourenço, Sérgio Ahrens, Susete do Rocio C. Penteado.

**Revisão gramatical:** Elly Claire Jansson Lopes

**Normalização:** Lidia Woronkoff

---

RESENDE, M.D.V. de. Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes. Colombo: Embrapa Florestas, 2000. 101p. (*Embrapa Florestas*. Documentos, 47).

ISSN 1517-536X

1. Estatística. 2. Modelos mistos. 3. Planta perene. I. Título.  
II. Série.

CDD 519.5

---

©*Embrapa*, 2000

Produção:

ÁREA DE COMUNICAÇÕES E NEGÓCIOS

Supervisor: Miguel Haliski

LAYOUT DA CAPA:

Cleide da S.N.F. de Oliveira

DIAGRAMAÇÃO

Marta de Fátima Vencato

IMPRESSÃO

Gráfica Radial - Fone: 333-9593

Dezembro/2000

# Sumário

1 INTRODUÇÃO .....	7
2 MODELOS LINEARES MISTOS, BLUP E REML .....	8
3 PROGRAMAS COMPUTACIONAIS .....	12
4 PREDIÇÃO DE VALORES GENÉTICOS INTRAPOPULACIONAIS (INTRAESPECÍFICOS) .....	14
4.1 DELINEAMENTO EM BLOCOS AO ACASO, PROGÊNIES DE POLINIZAÇÃO ABERTA, UMA SÓ POPULAÇÃO .....	14
4.1.1 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO, UM SÓ CARÁTER E UMA SÓ POPULAÇÃO .....	14
4.1.2 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO, UM SÓ CARÁTER E UMA SÓ POPULAÇÃO .....	17
4.1.3 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E UM SÓ CARÁTER, AVALIADO EM VÁRIOS LOCAIS (EXPERIMENTOS) COM ALGUMAS PROGÊNIES OU TRATAMENTOS COMUNS .....	20
4.1.4 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO E UM SÓ CARÁTER, AVALIADO EM VÁRIOS EXPERIMENTOS COM ALGUMAS PROGÊNIES OU TRATAMENTOS COMUNS .....	25
4.2 DELINEAMENTO EM BLOCOS AO ACASO, PROGÊNIES DE POLINIZAÇÃO CONTROLADA, UMA SÓ POPULAÇÃO .....	27
4.2.1 AVALIAÇÃO DE PROGÊNIES DE IRMÃOS GERMANOS OBTIDAS SOB CRUZAMENTOS DIALÉLICOS, FATORIAIS OU HIERÁRQUICOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO, UM SÓ CARÁTER E UMA SÓ POPULAÇÃO .....	27
4.2.2 AVALIAÇÃO DE PROGÊNIES DE IRMÃOS GERMANOS OBTIDAS SOB CRUZAMENTOS DIALÉLICOS, FATORIAIS OU HIERÁRQUICOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO, UM SÓ CARÁTER E UMA SÓ POPULAÇÃO .....	30
4.2.3 AVALIAÇÃO DE PROGÊNIES DE IRMÃOS GERMANOS OBTIDAS SOB CRUZAMENTOS DIALÉLICOS, FATORIAIS OU HIERÁRQUICOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E UM SÓ CARÁTER AVALIADO EM VÁRIOS EXPERIMENTOS, COM TRATAMENTOS COMUNS .....	32

<b>4.2.4 AVALIAÇÃO DE PROGÊNIES DE IRMÃOS GERMANOS OBTIDAS SOB CRUZAMENTOS DIALÉLICOS, FATORIAIS OU HIERÁRQUICOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO, UM SÓ CARÁTER, UMA SÓ POPULAÇÃO, AVALIADA EM VÁRIOS EXPERIMENTOS COM PROGÊNIES OU TRATAMENTOS COMUNS .....</b>	<b>34</b>
<b>4.3 DELINEAMENTO EM BLOCOS AO ACASO, PROGÊNIES DE POLINIZAÇÃO ABERTA, VÁRIAS POPULAÇÕES .....</b>	<b>35</b>
<b>4.3.1 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS (POLINIZAÇÃO ABERTA) DE VÁRIAS POPULAÇÕES (PROCEDÊNCIAS), NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E UM SÓ CARÁTER .....</b>	<b>35</b>
<b>4.3.2 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS (POLINIZAÇÃO ABERTA) DE VÁRIAS POPULAÇÕES (PROCEDÊNCIAS), NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO E UM SÓ CARÁTER .....</b>	<b>39</b>
<b>4.3.3 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS (POLINIZAÇÃO ABERTA) DE VÁRIAS POPULAÇÕES (PROCEDÊNCIAS), NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E UM SÓ CARÁTER, AVALIADO EM VÁRIOS EXPERIMENTOS COM ALGUMAS PROGÊNIES OU TRATAMENTOS COMUNS .....</b>	<b>43</b>
<b>4.3.4 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS (POLINIZAÇÃO ABERTA) DE VÁRIAS POPULAÇÕES (PROCEDÊNCIAS), NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO E UM SÓ CARÁTER, AVALIADO EM VÁRIOS EXPERIMENTOS COM ALGUMAS PROGÊNIES OU TRATAMENTOS COMUNS .....</b>	<b>46</b>
<b>4.4 DELINEAMENTO EM LÁTICE, PROGÊNIES DE POLINIZAÇÃO ABERTA, UMA POPULAÇÃO</b>	<b>50</b>
<b>4.4.1 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM LÁTICE, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO, UM SÓ CARÁTER E UMA SÓ POPULAÇÃO .....</b>	<b>50</b>
<b>4.4.2 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM LÁTICE, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO, UM SÓ CARÁTER E UMA SÓ POPULAÇÃO .....</b>	<b>52</b>
<b>4.4.3 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM LÁTICE, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E UM SÓ CARÁTER, AVALIADO EM VÁRIOS EXPERIMENTOS COM ALGUMAS PROGÊNIES OU TRATAMENTOS COMUNS .....</b>	<b>55</b>
<b>4.4.4 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM LÁTICE, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO E UM SÓ CARÁTER, AVALIADO EM VÁRIOS EXPERIMENTOS COM ALGUMAS PROGÊNIES OU TRATAMENTOS COMUNS .....</b>	<b>57</b>

<b>4.5 DELINEAMENTO EM LÁTICE, PROGÊNIES DE POLINIZAÇÃO ABERTA, VÁRIAS POPULAÇÕES .....</b>	<b>59</b>
<b>4.5.1 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS (POLINIZAÇÃO ABERTA) DE VÁRIAS POPULAÇÕES (PROCEDÊNCIAS), NO DELINEAMENTO EM LÁTICE, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E UM SÓ CARÁTER .....</b>	<b>59</b>
<b>4.5.2 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS (POLINIZAÇÃO ABERTA) DE VÁRIAS POPULAÇÕES (PROCEDÊNCIAS), NO DELINEAMENTO EM LÁTICE, COM VÁRIAS PLANTAS POR PARCELA, VÁRIAS MEDIÇÕES POR INDIVÍDUO E UM SÓ CARÁTER .....</b>	<b>62</b>
<b>4.6 AVALIAÇÃO APENAS DE POPULAÇÕES .....</b>	<b>65</b>
<b>4.6.1 AVALIAÇÃO DE VÁRIAS POPULAÇÕES (PROCEDÊNCIAS), NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E UM SÓ CARÁTER .....</b>	<b>65</b>
<b>4.7 AVALIAÇÃO SIMULTÂNEA DE CARACTERES .....</b>	<b>68</b>
<b>4.7.1 AVALIAÇÃO DE PROGÊNIES DE MEIOS IRMÃOS, NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA, UMA MEDIÇÃO POR INDIVÍDUO E VÁRIOS CARACTERES, EM UM SÓ LOCAL .....</b>	<b>68</b>
<b>5 PREDIÇÃO DE VALORES GENÉTICOS E GENOTÍPICOS INTERPOPULACIONAIS (INTERESPECÍFICOS) .....</b>	<b>71</b>
<b>5.1 HÍBRIDOS ENVOLVENDO DUAS ESPÉCIES OU POPULAÇÕES .....</b>	<b>71</b>
<b>5.2 HÍBRIDOS ENVOLVENDO TRÊS OU MAIS ESPÉCIES .....</b>	<b>84</b>
<b>6 TESTES CLONALIS .....</b>	<b>85</b>
<b>6.1 AVALIAÇÃO DE CLONES NÃO APARENTADOS NO DELINEAMENTO EM BLOCOS AO ACASO COM VÁRIAS PLANTAS POR PARCELA (VÁLIDO TAMBÉM PARA LINHAGENS NÃO APARENTADAS DE ESPÉCIES AUTÓGAMAS) .....</b>	<b>85</b>
<b>6.2 AVALIAÇÃO DE CLONES APARENTADOS NO DELINEAMENTO EM BLOCOS AO ACASO, COM VÁRIAS PLANTAS POR PARCELA .....</b>	<b>87</b>
<b>7 MODELOS NÃO LINEARES PARA VARIÁVEIS BINOMIAIS E CATEGÓRICAS .....</b>	<b>89</b>
<b>8 AJUSTE DE COVARIÁVEL E ANÁLISE DE COVARIÂNCIA .....</b>	<b>96</b>
<b>9 ESPÉCIES COM SISTEMA REPRODUTIVO MISTO .....</b>	<b>97</b>
<b>10 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>98</b>



# **ANÁLISE ESTATÍSTICA DE MODELOS MISTOS VIA REML/BLUP NA EXPERIMENTAÇÃO EM MELHORAMENTO DE PLANTAS PERENES**

Marcos Deon Vilela de Resende

## **1. Introdução**

No melhoramento de plantas perenes, as técnicas de avaliação genética desempenham papel fundamental, pois permitem a predição dos valores genéticos aditivos e genotípicos dos candidatos a seleção, propiciando uma seleção mais acurada. Tais técnicas são relevantes tanto para o melhoramento intrapopulacional quanto interpopulacional visando à utilização de híbridos heteróticos.

As técnicas ótimas de avaliação genética envolvem, simultaneamente, a predição de valores genéticos e a estimativa de componentes de variância. De maneira genérica, o procedimento ótimo de predição de valores genéticos é o BLUP (melhor predição linear não viciada) ao nível individual (Henderson & Quaas, 1976). Para o caso balanceado, os preditores BLUP ao nível individual equivalem aos índices de seleção multi-efeitos (Resende & Higa, 1994a), os quais envolvem todos os efeitos aleatórios do modelo estatístico associado às observações fenotípicas. A predição usando BLUP ou os índices multi-efeitos assume que os componentes de variância são conhecidos. Entretanto, na prática são necessárias estimativas fidedignas dos componentes de variância (parâmetros genéticos) de forma a se obter o que se denomina BLUP empírico (Harville & Carriquiry, 1992). Atualmente, o procedimento padrão de estimativa de componentes de variância é o da máxima verossimilhança restrita (REML), desenvolvido por Patterson & Thompson (1971).

A estimativa de parâmetros genéticos associados à seleção no contexto do melhoramento de plantas anuais é bem descrito em várias obras publicadas no Brasil (Vencovsky, 1987; Vencovsky & Barriga, 1992; Ramalho et al., 1993; Cruz & Regazzi, 1994; Cruz, 1997; Ramalho et al., 2000). Por outro lado, a estimativa e predição no contexto do melhoramento de plantas perenes demanda o uso da metodologia de modelos mistos (REML/BLUP) ao nível individual, a qual vem sendo aplicada ao melhoramento de espécies florestais como o eucalipto, o pinus, a acácia-negra e a seringueira (Resende et al., 1993; 1996; Bueno Filho, 1997; Resende & Fernandes, 1999; Resende et al., 1998; Costa et al., 1999; Kalil et al., 2000), de espécies produtoras de alimentos estimulantes (contendo alcalóides como a cafeína e a teobromina) tais quais a erva-mate, o cacau e o café (Resende et al., 2000a; Resende & Dias, 2000;

Resende et al., 2000b), de fruteiras como a acerola (Paiva et al., 2000, Resende, 2000) e de palmáceas como a pupunha (Farias Neto & Resende, 2000), mas necessita ser difundida e disseminada. As técnicas de estimação baseadas no método de quadrados mínimos tais como a análise de variância não são as mais recomendadas para aplicação ao melhoramento de plantas perenes.

Este artigo tem como objetivo descrever, em termos práticos, os modelos mistos e suas estruturas de médias e variâncias, estimadores e preditores associados aos principais delineamentos experimentais e de cruzamentos empregados no melhoramento de plantas perenes. São contemplados os delineamentos experimentais de blocos ao acaso e látice, delineamentos de cruzamento em polinização aberta e controlada, medidas simples e repetidas, modelos univariados e multivariados, avaliação de progênies intrapopulacionais e híbridas, avaliação de clones, uma e várias populações, avaliação em experimentos simples e repetidos através dos locais (contemplando a interação genótipo x ambiente), avaliação simultânea de caracteres. Todas estas situações são sobrepostas gerando em torno de 30 modelos diferentes. São consideradas variáveis contínuas (modelos lineares) e variáveis discretas (modelos não lineares associados à técnica de modelos lineares generalizados). São também abordados com detalhes, aspectos referentes ao uso dos softwares ASREML (Gilmour et al., 2000) e DFREML (Meyer, 1998).

Todos os aspectos considerados neste trabalho são essenciais na prática do melhoramento de plantas perenes. Adicionalmente, são fundamentais aos trabalhos acadêmicos e técnico-científicos, referindo-se ao conteúdo da metodologia dos trabalhos científicos que forem desenvolvidos através do uso da abordagem de modelos lineares e não lineares mistos.

Acredita-se que este trabalho sirva de referência ao melhoramento de plantas perenes no Brasil tais quais: (i) espécies florestais: eucalipto, pinus, acácia-negra, grevílea, seringueira, leucena, etc; (ii) espécies produtoras de alimentos estimulantes: erva-mate, cacau, café, guaraná, chá-da-índia, etc; (iii) fruteiras: caju, acerola, cupuaçu, maçã, graviola, etc; (iv) palmáceas: coco, dendê, açaí, pupunha, juçara, tamareira, palmeira real, etc.

## 2. Modelos lineares mistos, BLUP e REML

Um modelo linear misto geral é da forma (Henderson, 1984):

$$y = Xb + Za + e \quad (1),$$

com as seguintes distribuições e estruturas de médias e variâncias:

$$a \sim N(0, G) \quad E(y) = Xb$$

$$e \sim N(0, R) \quad Var(y) = V = ZGZ' + R$$

em que:

y: vetor de observações;

b: vetor paramétrico dos efeitos fixos, com matriz de incidência X;

a: vetor paramétrico dos efeitos aleatórios, com matriz de incidência Z;

e: vetor de erros aleatórios;

G: matriz de variância – covariância dos efeitos aleatórios;

R: matriz de variância – covariância dos erros aleatórios;

O: vetor nulo.

Assumindo como conhecidos G e R, a simultânea estimação dos efeitos fixos e predição dos efeitos aleatórios pode ser obtida pelas equações de modelo misto dadas por:

$$\begin{bmatrix} Z'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

A solução deste sistema para  $\hat{b}$  e  $\hat{a}$  conduz a resultados idênticos aos obtidos por:

$\hat{b} = (X'V^{-1}X)^{-1}X'V^{-1}y$ : estimador de quadrados mínimos generalizados (GLS)  
ou melhor estimador linear não viciado (BLUE) de b;

$\hat{a} = GZ'V^{-1}(y - X\hat{b}) = CV^{-1}(y - X\hat{b})$ : melhor preditor linear não viciado (BLUP)  
de a; em que C = GZ' = matriz de covariância entre  
a e y.

Quando G e R não são conhecidas, os componentes de variância a eles associados podem ser estimados eficientemente empregando-se o procedimento REML (Patterson & Thompson, 1971; Searle et al., 1992). Exceto por uma constante, a função de verossimilhança restrita a ser maximizada, é dada por:

$$L = -\frac{1}{2} (\log|XH^{-1}X| + \log|H| + v \log \sigma_e^2 + y'Py/\sigma_e^2)$$

$$= -\frac{1}{2} (\log|C| + \log|R| + \log|G| + v \log \sigma_e^2 + y'Py/\sigma_e^2)$$

em que:

$$H = R + ZGZ'; \quad P = H^{-1} - H^{-1}X(X'H^{-1}X)^{-1}X'H^{-1};$$

$v = N - r(x)$  = graus de liberdade, em que  $N$  é o número total de dados e  $r(x)$  é o posto da matriz  $X$ ;

$C$  = matriz dos coeficientes das equações de modelo misto.

A função ( $L$ ) de verossimilhança restrita expressa em termos do logaritmo, pode ser maximizada (visando obter as estimativas REML dos componentes de variância) empregando-se diferentes algoritmos tais quais: (i) "Expectation – Maximization" (EM) de Dempster et al. (1977); (ii) "Derivative Free" (DF) de Graser et al. (1987); (iii) "Average Information" (AI) de Gilmour et al. (1995). Estes algoritmos geraram as denominações EM-REML, DF-REML e AI-REML. Dentre estes, o algoritmo EM é o mais acurado, mas também o mais lento. O algoritmo DF é rápido e acurado quando o número de componentes de variância não é muito grande. Para modelos complexos, o algoritmo AI é mais rápido e acurado do que o DF (Johnson & Thompson, 1995).

Sendo geral, o modelo (1) contempla vários modelos inerentes às diferentes situações, tais quais:

**(a) Modelo univariado, ajustando apenas o vetor de efeitos aditivos (a)**

$a$  : vetor de efeitos genéticos aditivos;

$$G = A\sigma_a^2; \quad R = I\sigma_e^2, \text{ em que:}$$

$\sigma_a^2$  : variância genética aditiva;

$A$  : matriz de correlação genética aditiva entre os indivíduos em avaliação;

$\sigma_e^2$  : variância residual.

**(b) Modelo univariado com medidas repetidas, ajustando os efeitos aditivos (a\*) e de ambiente permanente (p) (Modelo de Repetibilidade)**

$$y = Xb + Za + e \quad \text{e} \quad \text{Var}(a^*) = A\sigma_a^2; \quad \text{Var}(p) = I\sigma_p^2; \quad R = I\sigma_e^2$$

$= Xb + Z_1 a^* + Z_2 p + e$ , em que:

$\sigma_p^2$  = variância dos efeitos permanentes.

**(c) Modelo multivariado, ajustando os efeitos aditivos**

No caso bivariado tem-se:

$$Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}; \quad a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix};$$

$$G = A \otimes G_o; \quad R = I \otimes R_o;$$

$$G_o = \begin{bmatrix} \sigma_{a_{11}}^2 & \sigma_{a_{12}}^2 \\ \sigma_{a_{21}}^2 & \sigma_{a_{22}}^2 \end{bmatrix}; \quad R_o = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}}^2 \\ \sigma_{e_{21}}^2 & \sigma_{e_2}^2 \end{bmatrix} \quad \text{ou} \quad R_o = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}, \text{ em que:}$$

$\sigma_{a_{12}}$  = covariância genética aditiva entre os caracteres 1 e 2;

$\sigma_{e_{12}}$  = covariância ambiental entre os caracteres 1 e 2.

**(d) Modelo geoestatístico para análise espacial**

R = S : matriz não diagonal que considera a correlação entre resíduos, por exemplo, linhas auto-regressivas e colunas auto-regressivas, para contemplar a autocorrelação espacial entre as observações.

Os modelos geoestatísticos permitem estudar a variabilidade espacial do solo nas áreas experimentais, através do uso de procedimentos que permitem um melhor critério de estratificação ambiental (para seleção massal ou para melhor definição dos efeitos fixos no procedimento BLUP). Neste contexto, a análise espacial é realizada simultaneamente (Cullis et al., 1998) à predição BLUP. Conhecimentos e modelos de séries temporais são muito úteis nesta área de pesquisa. Também, muitas técnicas empregadas na área de geologia tais quais a construção de semivariogramas e a realização de krigagens são empregados na análise espacial.

De maneira genérica, as variâncias dos erros de estimação e predição dos efeitos fixos e aleatórios são dadas por  $Var \begin{bmatrix} \hat{b} & - & b \\ \hat{a} & - & a \end{bmatrix} = C^{-1}$ , para um modelo incluindo os efeitos fixos ( $b$ ) e aleatórios ( $a$ ), em que  $C^{-1}$  é a inversa da matriz dos coeficientes das equações de modelo misto.

A partir da variância do erro de predição (PEV) dos valores genéticos a acurácia é dada por  $r_{\hat{aa}_i} = [1 - PEV_i / \sigma_a^2]^{1/2}$ .

### **3. Programas computacionais**

A implementação computacional da metodologia de modelos mistos baseia-se fortemente em métodos numéricos, notadamente em álgebra linear numérica visando a obtenção da solução iterativa das equações de modelo misto (obtenção do BLUP) e no cálculo numérico para a maximização/minimização de funções de várias variáveis visando a obtenção das estimativas REML.

Os algoritmos para obtenção de estimativas REML podem ser agrupados de acordo com a ordem das derivadas usadas. Assim, tem-se (i) não derivativo (DF-REML), baseado em procura direta; (ii) baseado em derivadas parciais de primeira ordem (EM-REML); (iii) baseado em derivadas parciais de primeira e segunda ordens (AI-REML). O algoritmo AI é um procedimento derivativo melhorado, o qual fundamenta-se no uso dos métodos de Newton, que usam as derivadas primeira e segunda da função de verossimilhança. Tal algoritmo fundamenta-se na utilização da informação advinda da média das derivadas segundas observadas e esperadas da função de verossimilhança, de forma que o termo que contém os traços dos produtos da matriz inversa é cancelado, restando uma expressão mais simples para computação. Técnicas de matrizes esparsas são empregadas no cálculo dos elementos da inversa da matriz dos coeficientes, os quais são necessários para as derivadas primeiras da função de verossimilhança. Este algoritmo é também denominado Quasi-Newton (Gilmour et al., 1995), o qual aproxima a matriz Hessiana (matriz de derivadas segundas) pela média das informações observadas e esperadas. A informação observada é uma medida da curvatura da função (ou do seu log) de verossimilhança e a informação esperada é a própria informação de Fisher.

Os algoritmos DF ganharam popularidade devido as suas flexibilidades quanto aos modelos (Meyer, 1989; 1991) e vários softwares foram desenvolvidos, tais quais o DFREML (Meyer, 1988; 1998) e o MTDFREML (Boldman et al., 1995). Entretanto, as dificuldades de convergência em modelos mais complexos geraram um novo interesse em métodos baseados em primeira

e segunda derivadas da função de verossimilhança. Assim, o algoritmo AI foi incorporado ao DFREML (Meyer, 1998) e desenvolveu-se o software ASREML (Gilmour et al., 2000). Quando o número de parâmetros a serem estimados é pequeno, os algoritmos DF são vantajosos computacionalmente. Por outro lado, quando o número de parâmetros é grande os algoritmos EM são mais eficientes que o DF. Nesta mesma situação, o algoritmo AI supera o DF e o EM.

Os três softwares mencionados (ASREML, DFREML e MTDFREML) são os mais utilizados no melhoramento de plantas no Brasil e permitem a análise de arquivos com número de dados superior a 100.000. Por incorporarem o algoritmo AI e também a técnica de regressão aleatória, os softwares DFREML e ASREML têm sido os mais utilizados e eficientes. O software ASREML contempla, adicionalmente, a técnica de análise espacial associada ao BLUP e permite também a análise de modelos não lineares para variáveis binomiais, empregando a técnica de modelos lineares generalizados. Este software permite, praticamente, a análise de qualquer modelo, mesmo os mais complexos, e dificilmente será superado.

Neste artigo, são apresentados aspectos da utilização do software DFREML (Meyer, 1998). Os programas do DFREML (Versão 3.0b) foram escritos em Fortran 90 (as linguagens de programação mais recomendadas para o desenvolvimento de softwares na área de componentes de variância são Fortran 90 e C++) e possuem duas versões específicas: Unix e PC (DOS). Quatro programas constituem o DFREML: DFPREP que se destina à recodificação dos efeitos fixos e aleatórios na ordem de processamento e à construção da matriz de parentesco; DFUNI destinado a análises univariadas permitindo o ajuste de vários efeitos aleatórios não correlacionados, adicionais aos efeitos aditivos; DXMUX destinado a análise multivariada, mas permitindo também o ajuste de diferentes modelos para os caracteres individuais, os quais podem diferir tanto nos efeitos fixos quanto aleatórios. Este programa permite o ajuste de, no máximo, dois efeitos aleatórios adicionais não correlacionados; DXMRR: destinado a análise de medidas repetidas (dados longitudinais), estimando funções de covariância e regressões aleatórias.

Os arquivos de dados ou de pedigree devem possuir a extensão .PRN ou .DAT ou .TXT. O arquivo de resultados mais importantes do DFREML é o DF66#DAT, o qual apresenta as estimativas dos componentes de variância e seus desvios padrões, o Log L e as soluções para os efeitos fixos e aleatórios (valores genéticos preditos).

Nos tópicos seguintes são apresentadas as formas (seqüências de colunas nos arquivos de dados) dos arquivos para vários modelos empregados no melhoramento de plantas perenes.

## **4. Predição de valores genéticos intrapopulacionais (intraespecíficos)**

Os valores genéticos aditivos intrapopulacionais preditos são úteis tanto para o melhoramento intrapopulacional (intraespecífico) quanto para o melhoramento interpopulacional (interespecífico). Para o melhoramento intrapopulacional, norteia a seleção visando ao melhoramento progressivo da espécie. Para o melhoramento interpopulacional, permite a escolha criteriosa dos indivíduos a serem empregados nos cruzamentos, bem como conduz ao próprio melhoramento do híbrido, em função dos melhoramentos realizados nas espécies puras. A seguir, são apresentados alguns modelos de avaliação genética (estimação de componentes de variância e predição de valores genéticos) em espécies perenes. São apresentadas, também, as formas de organização das colunas de dados visando à utilização do software DFREML (Meyer, 1998).

### **4.1 Delineamento em blocos ao acaso, progênies de polinização aberta, uma só população**

#### **4.1.1 Avaliação de progênies de meios irmãos, no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo, um só caráter e uma só população**

##### **Modelo linear misto (modelo aditivo univariado)**

$$y = Xb + Za + Wc + e, \text{ em que}$$

y, b, a, c e e: vetores de dados, dos efeitos de blocos (fixos), dos efeitos genéticos aditivos (aleatórios), de efeitos de parcela (aleatórios) e dos erros aleatórios, respectivamente.

X, Z e W: matrizes de incidência para b, a e c, respectivamente.

##### **Distribuições e estruturas de médias e variâncias**

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

$$a|A, \sigma_a^2 \sim N(0, A \sigma_a^2)$$

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

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

$$\text{Cov}(a, c') = 0; \quad \text{Cov}(a, e') = 0; \quad \text{Cov}(c, e') = 0$$

ou seja:

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

$$G = A \sigma_a^2$$

$$R = I \sigma_c^2$$

$$C = I \sigma_e^2$$

$$V = ZA \sigma_a^2 Z' + WI \sigma_c^2 W' + I \sigma_e^2 = ZGZ' + WCW' + R.$$

### Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W \\ W'X & W'Z & W'W + I\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2-c^2}{h^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1-h^2-c^2}{c^2}$$

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2} : \text{herdabilidade individual no sentido restrito no bloco;}$$

$c^2 = \sigma_c^2 / (\sigma_a^2 + \sigma_c^2 + \sigma_e^2)$  : correlação devida ao ambiente comum da parcela;

$\sigma_a^2$ : variância genética aditiva;

$\sigma_c^2$ : variância entre parcelas;

$\sigma_e^2$ : variância residual (ambiental dentro de parcelas + não aditiva);

A : matriz de correlação genética aditiva entre os indivíduos em avaliação.

As soluções para as equações de modelo misto devem ser obtidas por métodos iterativos de resolução de sistemas de equações lineares, tais como o método de Gauss Seidel. A solução direta via inversão da matriz dos coeficientes é impossível na prática.

Para o caso balanceado, as soluções obtidas para  $\hat{a}$ , a partir das equações de modelo misto, eqüivalem às predições obtidas pelo índice multi-efeitos (Resende & Higa, 1994a), dado por:

$I = b_1(Y_{ijk} - \bar{Y}_{ij.}) + b_2(\bar{Y}_{i..} - \bar{Y}_{...}) + b_3(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$ , em que:

$$b_1 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_{dp}^2}; \quad b_2 = \frac{\left[1+(nb-1)\rho_a\right] \sigma_a^2}{\sigma_{fa}^2 + \sigma_c^2/b + \sigma_{dp}^2/nb}; \quad b_3 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_{dp}^2/n + \sigma_c^2}$$

$\rho_a$ : correlação genética aditiva intraclasse ( $\rho_a = 0,25$  para famílias de meios irmãos);

$\sigma_{fa}^2$ ,  $\sigma_c^2$  e  $\sigma_{dp}^2$ : variância entre famílias, entre parcelas e dentro de parcelas, respectivamente.

$Y_{ijk}$ ,  $\bar{Y}_{ij.}$ ,  $\bar{Y}_{i..}$ ,  $\bar{Y}_{.j.}$  e  $\bar{Y}_{...}$ : valor fenotípico individual, média da parcela, média da progénie, média do bloco e média geral, respectivamente.

Assim, para o caso balanceado, este índice é BLUP, ao passo que as formas de seleção combinada usando o valor individual como desvio da média da parcela (Bueno Filho, 1992; Resende & Higa, 1994b) ou como desvio da média do bloco (Pires et al., 1996) não são BLUP (Resende & Fernandes, 1999).

### **Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y' y - \hat{b}' X' y - \hat{a}' Z' y - \hat{c}' W' y]/[N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \text{tr } (A^{-1} C^{22})]/q$$

$$\hat{\sigma}_c^2 = [\hat{c}' c + \hat{\sigma}_e^2 \text{tr } C^{33}]/s, \text{ em que:}$$

$C^{22}$  e  $C^{33}$  advém de:

$$C^{-1} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}^{-1} = \begin{bmatrix} C^{11} & C^{12} & C^{13} \\ C^{21} & C^{22} & C^{23} \\ C^{31} & C^{32} & C^{33} \end{bmatrix}$$

C: matriz dos coeficientes das equações de modelo misto;

tr: operador traço matricial;

r(x): posto da matriz X;

N, q, s: número total de dados, número de indivíduos e número de parcelas, respectivamente.

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

Indivíduo	Pai	Mãe	Bloco	Parcela	Variável 1	...	Variável n
-----------	-----	-----	-------	---------	------------	-----	------------

Neste caso, é necessário apenas um arquivo, o qual funciona ao mesmo tempo como arquivo de pedigree e como arquivo de dados. Deve ser executado os subprograma DFPREP e em seguida o DFUNI.

Os estimadores e preditores apresentados podem ser utilizados com eficiência em testes de progênies de irmãos germanos (obtidas sob o delineamento de cruzamento em pares simples) desde que a dominância do caráter seja baixa.

#### **4.1.2 Avaliação de progênies de meios irmãos, no delineamento em blocos ao acaso, com várias plantas por parcela, várias medições por indivíduo, um só caráter e uma só população**

##### **Modelo linear misto (modelo aditivo univariado, de repetibilidade)**

$y = Xb + Za + Wc + Tp + e$ , em que

p: vetor de efeitos permanentes (ambiente permanente dentro de parcela + efeitos genéticos não aditivos);

T: matriz de incidência para p.

Neste modelo, os efeitos fixos de blocos ( $b-1$  graus de liberdade), medições ( $m-1$  graus de liberdade) e interação medição x bloco [ $(b-1)(m-1)$  graus de liberdade], podem ser ajustados em um único efeito (denominado combinação bloco-medição com  $mb$  elementos ou níveis e  $mb-1$  graus de liberdade), procedimento este que é estatisticamente correto e computacionalmente

desejável e necessário. Os efeitos temporários da interação progêneres x medições e progêneres x medições x blocos são incorporados ao vetor  $\mathbf{e}$ , em conjunto com o efeito de ambiente temporário propriamente dito. Os efeitos  $c$  de parcela, neste modelo, referem-se ao ambiente permanente entre parcelas.

### Distribuições e estruturas de médias e variâncias

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

$$a|A, \sigma_a^2 \sim N(0, \sigma_a^2)$$

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

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

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

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

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

ou seja:

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

$$P = I \sigma_p^2$$

$$V = ZA \sigma_a^2 Z' + WI \sigma_c^2 W' + TI \sigma_p^2 T' + I \sigma_e^2.$$

## Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'T \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'T \\ W'X & W'Z & W'W + I\lambda_2 & W'T \\ T'X & T'Z & T'W & T'T + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{1-\rho}{h^2} = \frac{\sigma_e^2}{\sigma_a^2}; \quad \lambda_2 = \frac{1-\rho}{c^2} = \frac{\sigma_e^2}{\sigma_c^2}; \quad \lambda_3 = \frac{1-\rho}{p^2} = \frac{\sigma_e^2}{\sigma_p^2}.$$

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2}$  : herdabilidade individual no sentido restrito no bloco,  
em uma dada medição;

$\rho = \frac{\sigma_a^2 + \sigma_c^2 + \sigma_p^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2}$  : repetibilidade individual no bloco;

$p^2 = \frac{\sigma_p^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2}$  : coeficiente de determinação dos efeitos permanentes  
dentro da parcela;

$c^2 = \frac{\sigma_c^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2}$  : correlação devida ao ambiente comum da parcela.

## Estimadores iterativos dos componentes de variância por REML via algoritmo EM

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}' Z'y - \hat{c}' W'y - \hat{p}' T'y] / [N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \operatorname{tr}(A^{-1} C^{22})] / q$$

$$\hat{\sigma}_c^2 = [\hat{c}' c + \hat{\sigma}_e^2 \operatorname{tr} C^{33}] / s$$

$$\hat{\sigma}_p^2 = [\hat{p}' \hat{p} + \hat{\sigma}_e^2 \operatorname{tr} C^{44}] / q, \text{ em que:}$$

$C^{22}$ ,  $C^{33}$  e  $C^{44}$  advém de:

$$C^{-1} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{bmatrix}^{-1} = \begin{bmatrix} C^{11} & C^{12} & C^{13} & C^{14} \\ C^{21} & C^{22} & C^{23} & C^{24} \\ C^{31} & C^{32} & C^{33} & C^{34} \\ C^{41} & C^{42} & C^{43} & C^{44} \end{bmatrix}$$

### Seqüência de colunas no arquivo de dados para análise no software DFREML

Indivíduo    Pai    Mãe    Bloco-Medição    Parcela    Permanente    Variável 1    . . .    Variável n

É necessário apenas um arquivo, o qual funciona simultaneamente como arquivo de dados e como arquivo de pedigree. Devem ser executados sequencialmente os subprogramas DFPREP e DFUNI.

#### **4.1.3 Avaliação de progêneros de meios irmãos, no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo e um só caráter, avaliado em vários locais (experimentos) com algumas progêneros ou tratamentos comuns**

##### **Modelo linear misto (modelo aditivo multivariado)**

$y = Xb + Za + Wc + e$ , em que:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_\ell \end{bmatrix} = \begin{bmatrix} X_1 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_\ell \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_\ell \end{bmatrix} + \begin{bmatrix} Z_1 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_\ell \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_\ell \end{bmatrix} + \begin{bmatrix} W_1 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_\ell \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\ell \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_\ell \end{bmatrix}$$

Este modelo multivariado trata um mesmo caráter em diferentes locais como sendo diferentes caracteres (Resende et al., 1999).

## Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

$$Var(y_1) = V_1 = Z_1 A \sigma_{a_1}^2 Z_1' + W_1 I \sigma_{c_1}^2 W_1' + I \sigma_{e_1}^2$$

$$Var(y_2) = V_2 = Z_2 A \sigma_{a_2}^2 Z_2' + W_2 I \sigma_{c_2}^2 W_2' + I \sigma_{e_2}^2$$

Para o caso envolvendo três experimentos, tem-se:

$$Var \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ c_1 \\ c_2 \\ c_3 \\ e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & A\sigma_{a_{13}} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & A\sigma_{a_{23}} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{13}} & A\sigma_{a_{23}} & A\sigma_{a_3}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{c_3}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_3}^2 \end{bmatrix}$$

em que:

$\sigma_{a_1}^2$ ,  $\sigma_{a_2}^2$  e  $\sigma_{a_3}^2$ : variâncias genéticas aditivas, nos locais (ou experimentos) 1, 2 e 3, respectivamente;

$\sigma_{a_{12}}$ ,  $\sigma_{a_{13}}$  e  $\sigma_{a_{23}}$ : covariâncias genéticas aditivas, envolvendo as combinações de locais 1-2, 1-3 e 2-3, respectivamente, ou variâncias genéticas aditivas livres das interações genótipos x ambientes;

$\sigma_{c_1}^2$ ,  $\sigma_{c_2}^2$  e  $\sigma_{c_3}^2$ : variâncias entre parcelas, nos locais 1, 2 e 3, respectivamente;

$\sigma_{e_1}^2$ ,  $\sigma_{e_2}^2$  e  $\sigma_{e_3}^2$ : variâncias residuais, nos locais 1, 2 e 3, respectivamente.

### Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \end{bmatrix}, \text{ em que:}$$

$$X = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix}; \quad Z = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix}; \quad W = \begin{bmatrix} W_1 & 0 & 0 \\ 0 & W_2 & 0 \\ 0 & 0 & W_3 \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}; \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix}; \quad \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_3 \end{bmatrix}$$

$$R^{-1} = R_O^{-1} \otimes I; \quad G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C^{-1} = C_O^{-1} \otimes I$$

$$R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 & 0 \\ 0 & \sigma_{e_2}^2 & 0 \\ 0 & 0 & \sigma_{e_3}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} & \sigma_{a_{13}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 & \sigma_{a_{23}} \\ \sigma_{a_{13}} & \sigma_{a_{23}} & \sigma_{a_3}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c_1}^2 & 0 & 0 \\ 0 & \sigma_{c_2}^2 & 0 \\ 0 & 0 & \sigma_{c_3}^2 \end{bmatrix}$$

Os componentes de variância estão associados aos parâmetros  $h^2$ ,  $c^2$  e  $r_a$ , da seguinte maneira:

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{c_i}^2 = c_i^2 \sigma_{y_i}^2; \quad \sigma_{e_i}^2 = (1 - h_i^2 - c_i^2) \sigma_{y_i}^2; \quad \sigma_{a_{ij}} = \rho_{a_{ij}} \sigma_{a_i} \sigma_{a_j}, \quad \text{em}$$

que:

$$\rho_{a_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i} \sigma_{a_j}} = \text{correlação genética entre o desempenho nos locais i e j};$$

$$\sigma_{y_i}^2 = \text{variância fenotípica ao nível de indivíduo no local i.}$$

A variância da interação genótipo x ambiente, para o caso balanceado, é dada por  $\sigma_{ae}^2 = \frac{1}{2} (\sigma_{a_i} - \sigma_{a_j})^2 + (1 - \rho_{a_{ij}}) \sigma_{a_i} \sigma_{a_j}$ .

Modelos multivariados e equações de modelo misto deste tipo podem também ser usados na seguintes situações: (i) avaliação de indivíduos de diferentes gerações, em diferentes locais (para indivíduos de diferentes gerações avaliados em um mesmo local, basta ajustar adequadamente os efeitos fixos de blocos - anos e usar o modelo univariado); (ii) avaliação de indivíduos em diferentes estágios (juvenil e adulto), cada estágio em diferentes experimentos.

### **Estimadores dos componentes de variância por REML**

Nesta situação, devido à complexidade do modelo e ao elevado número de componentes de variância a serem estimados, os algoritmos recomendados são o DF e o AI (preferencialmente este). Neste caso, a função geral a ser maximizada é aquela descrita no item 2.

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

**Experimento Indivíduo Pai Mãe Bloco Parcela Variável 1 . . . Variável n**

Neste caso, são necessários dois arquivos: um de dados, conforme estrutura apresentada acima e outro de pedigree formado pelas colunas **Indivíduo Pai Mãe**. Deve ser executado o subprogramma DFPREP e em seguida o DXMUX.

## Análises alternativas

A metodologia de análise apresentada anteriormente permite a seleção de indivíduos para cada ambiente específico, porém, usando também a informação de suas famílias em outros ambientes. Uma outra alternativa é a seleção de indivíduos visando ao plantio dos mesmos nos vários ambientes. Nesta situação, a seleção deve ser baseada no comportamento médio ao longo dos ambientes, inferido pelos valores genéticos médios dados por:  $a_1^* = (a_{11} + a_{12})/2$  e  $a_2^* = (a_{22} + a_{21})/2$  para os indivíduos avaliados nos locais 1 e 2, respectivamente. Neste caso, a variância genética aditiva estimada equivale a  $\text{Var} [(a_{11} + a_{12} + a_{22} + a_{21})/2] = (1/4)(\sigma_{a_1}^2 + \sigma_{a_{12}}^2 + \sigma_{a_2}^2 + \sigma_{a_{12}}^2) = (1/4)(\sigma_{a_1}^2 + \sigma_{a_2}^2 + 2\sigma_{a_{12}}^2)$ . Neste caso, o modelo é dado por  $y^* = Xb + Za^* + Wc^* + e^*$ , e os estimadores e preditores equivalem àqueles apresentados no item 4.1.1, sendo que os arquivos devem ser montados conforme uma estrutura univariada.

Outra situação refere-se à avaliação da eficiência da seleção indireta ou mesmo à prática da seleção de indivíduos mais estáveis. Neste caso, a seleção deve ser baseada no valor genético indireto predito e a variância genética aditiva estimada equivale à variância livre da interação genótipo x ambiente. Nesta situação, o modelo é dado por  $y^{**} = Xb + Za^{**} + Wc^{**} + Sf^{**} + e^{**}$ , e os estimadores e preditores equivalem àqueles apresentados no item 4.1.2, porém substituindo-se  $T_p$  por  $Sf^{**}$  e  $\sigma_p^2$  por  $\sigma_{f^{**}}^2$ .

Neste caso,  $\sigma_{f^{**}}^2$  estima  $(1/4)$  de  $\sigma_{ae}^2$  ao passo que  $\sigma_{a^{**}}^2$  equivale a  $\sigma_a^2$  ou seja, a variância genética aditiva livre da interação genótipo x ambiente.

Considerando os três modelos alternativos apresentados neste tópico, tem-se as seguintes equivalências, para o caso balanceado:

$$(i) \quad \sigma_a^2 = \sigma_{a^{**}}^2;$$

$$(ii) \quad \sigma_{ae}^2 = \frac{1}{2}(\sigma_{a_1} - \sigma_{a_2}) + (1 - \rho_{a_{12}}) \sigma_{a_1} \sigma_{a_2} = 4\sigma_{f^{**}}^2;$$

$$(iii) \quad \sigma_a^2 + \sigma_{ae}^2 = (\sigma_{a_1}^2 + \sigma_{a_2}^2)/2;$$

$$(iv) \quad \sigma_{a^{**}}^2 = \frac{2\sigma_a^2 + \sigma_{a_1}^2 + \sigma_{a_2}^2}{4} = \sigma_a^2 + (1/2) \sigma_{ae}^2 = \sigma_{a^{**}}^2 + (1/2) \sigma_{ae}^2;$$

$$(v) \quad \sigma_{c^{**}}^2 = (\sigma_{c_1}^2 + \sigma_{c_2}^2)/2;$$

$$(vi) \quad \sigma_{e^{**}}^2 = (\sigma_{e_1}^2 + \sigma_{e_2}^2)/2;$$

$$(vii) \quad \sigma_{c^{*}}^2 = \sigma_{c^{**}}^2 + (1/2) \sigma_{f^{**}}^2;$$

$$(viii) \quad \sigma_{e^{*}}^2 = \sigma_{e^{**}}^2.$$

Em resumo, tem-se as seguintes opções de modelo de acordo com os respectivos objetivos:

- (i) seleção de diferentes materiais genéticos para plantios em ambientes específicos: modelo  $y$ ;
- (ii) seleção de materiais genéticos para plantio em locais em que não foram testados: modelo  $y^{**}$ ;
- (iii) seleção de um mesmo material genético para plantio nos vários locais em que foram conduzidos os experimentos: modelo  $y^*$ .

Em termos computacionais, é mais indicado ajustar inicialmente o modelo  $y^{**}$  e verificar a significância da interação genótipo x ambiente. Caso a interação não seja significativa, pode-se adotar o modelo  $y^*$ , o qual conduzirá praticamente ao mesmo resultado. Também, neste caso, se a interação for não significativa,  $\rho_{a_{12}}$  tenderá a 1 e, portanto, os modelos  $y^*$  e  $y$  também conduzirão praticamente ao mesmo resultado.

#### **4.1.4 Avaliação de progêneres de meios irmãos, no delineamento em blocos ao acaso, com várias plantas por parcela, várias medições por indivíduo e um só caráter, avaliado em vários experimentos com algumas progêneres ou tratamentos comuns**

##### **Modelo linear misto (modelo aditivo multivariado, de repetibilidade)**

Considerando o caso bivariado, tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

## Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_1}^2 & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{p_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{p_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}, \quad \text{em}$$

que:

$\sigma_{p_1}^2$  e  $\sigma_{p_2}^2$ : variância permanente dentro de parcelas (ambiental + genética não aditiva) nos locais 1 e 2, respectivamente.

## Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}T \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}T \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}T \\ T'R^{-1}X & T'R^{-1}Z & T'R^{-1}W & T'R^{-1}T + P^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ T'R^{-1}y \end{bmatrix}, \quad \text{em}$$

que:

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}; \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix}; \quad \hat{p} = \begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \end{bmatrix}$$

$$R^{-1} = R_O^{-1} \otimes I; \quad G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C^{-1} = C_O^{-1} \otimes I; \quad P^{-1} = P_O^{-1} \otimes I$$

$$R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_2}^2 \\ \sigma_{a_1}^2 & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad P_O = \begin{bmatrix} \sigma_{p_1}^2 & 0 \\ 0 & \sigma_{p_2}^2 \end{bmatrix}$$

Os componentes de variância estão associados aos parâmetros  $h^2$ ,  $c^2$ ,  $r$  e  $r_a$ , da seguinte maneira:

$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2$ ;  $\sigma_{e_i}^2 = (1 - \rho_i) \sigma_{y_i}^2$ ;  $\sigma_{c_i}^2 = c_i^2 \sigma_{y_i}^2$ ;  $\sigma_{p_i}^2 = (\rho_i - h_i^2 - c_i^2) \sigma_{y_i}^2$ ;  $\sigma_{a_j} = \rho_{a_j} \sigma_{a_i} \sigma_{a_j}$ , em que  $r_i$  é a repetibilidade no local i.

### Estimadores dos componentes de variância por REML

Nesta situação, são recomendados os algoritmos AI (preferencialmente este) e DF. A função geral a ser maximizada é aquela descrita no item 2.

#### Seqüência de colunas no arquivo de dados para análise no software DFREML

Experimento Indivíduo Pai Mãe Bloco-Medição Parcela Permanente Variável 1 . . . Variável n

São necessários dois arquivos, o de dados e um de pedigree, formado pelas colunas Indivíduo Pai Mãe. Devem ser executados os subprogramas DFPREP e DXMUX.

## 4.2 Delineamento em blocos ao acaso, progênie de polinização controlada, uma só população

### 4.2.1 Avaliação de progênie de irmãos germanos obtidas sob cruzamentos dialélicos, fatoriais ou hierárquicos, no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo, um só caráter e uma só população

#### Modelo linear misto (modelo aditivo-dominante univariado)

$y = Xb + Za + Zd + Wc + e$ , em que

d: vetor aleatório dos efeitos de dominância.

## Estruturas de médias e variâncias

$$E \begin{bmatrix} y \\ a \\ d \\ c \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \text{Var} \begin{bmatrix} y \\ a \\ d \\ c \\ e \end{bmatrix} = \begin{bmatrix} V & ZA\sigma_a^2 & ZD\sigma_d^2 & WI\sigma_c^2 & I\sigma_e^2 \\ A\sigma_a^2 Z' & A\sigma_a^2 & 0 & 0 & 0 \\ D\sigma_d^2 Z' & 0 & D\sigma_d^2 & 0 & 0 \\ I\sigma_c^2 W' & 0 & 0 & I\sigma_c^2 & 0 \\ I\sigma_e^2 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}, \text{ em que:}$$

$$V = ZA\sigma_a^2 Z' + ZD\sigma_d^2 Z' + WI\sigma_c^2 W' + I\sigma_e^2.$$

## Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'Z & Z'W \\ Z'X & Z'Z & Z'Z + D^{-1}\lambda_2 & Z'W \\ W'X & W'Z & W'Z & W'W + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{d} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h_a^2 - c^2}{h^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_d^2} = \frac{1 - h_a^2 - c^2}{h_a^2 - h^2}; \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1 - h_a^2 - c^2}{c^2}.$$

$\sigma_d^2$  e  $h_a^2$ : variância genética de dominância e herdabilidade individual no sentido amplo, respectivamente;

D : matriz de correlação genética de dominância entre os indivíduos em avaliação.

O sistema apresentado prediz isoladamente os efeitos aditivos ( $\hat{a}$ ) e de dominância ( $\hat{d}$ ). Os valores genotípicos totais, dados por  $\hat{g} = \hat{a} + \hat{d}$ , podem ser preditos diretamente pelas equações de modelo misto:

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + G^{-1}\sigma_e^2 & Z'W \\ W'X & W'Z & W'W + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$G = A\sigma_a^2 + D\sigma_d^2$$

## **Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y' y - \hat{b}' X' y - \hat{a}' Z' y - \hat{d}' Z' y - \hat{c}' W' y] / [N - r(x)];$$

$$\hat{\sigma}_c^2 = [\hat{c}' c + \hat{\sigma}_e^2 \operatorname{tr} C^{44}] / s;$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \operatorname{tr} (A^{-1} C^{22})] / q;$$

$$\hat{\sigma}_d^2 = [\hat{d}' D^{-1} \hat{d} + \hat{\sigma}_e^2 \operatorname{tr} (D^{-1} C^{33})] / q.$$

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

<b>Indivíduo</b>	<b>Pai</b>	<b>Mãe</b>	<b>Bloco</b>	<b>Parcela</b>	<b>Variável 1</b>	<b>...</b>	<b>Variável n</b>
------------------	------------	------------	--------------	----------------	-------------------	------------	-------------------

Para ajuste dos efeitos de dominância, os mesmos devem ser considerados como “segundo efeito aleatório por indivíduo”. Neste caso, além dos arquivos de dados e de pedigree, deve ser fornecido um arquivo adicional com nome padrão DF45#DAT, referente à inversa da matriz de parentesco de dominância. Este arquivo não formatado deve fornecer todos os elementos não zero do triângulo inferior da matriz inversa e deve conter três colunas: um código inteiro (de 1 ao número de indivíduos na análise) referente ao número da coluna na matriz; um código inteiro referente ao número da linha (maior ou igual ao número da coluna) na matriz (de 1 ao número de indivíduos na análise); uma variável verdadeira (real) fornecendo o elemento da inversa da matriz de parentesco de dominância.

Para a estimação da variância genética de dominância, outra alternativa pode ser empregada, a qual não requer o fornecimento dos elementos da inversa da matriz de parentesco de dominância. Isto pode ser feito ajustando o efeito de família de irmãos germanos como um efeito aleatório adicional não correlacionado que pode ser designado efeito de dominância comum a uma família de irmãos germanos. O vetor de soluções para este efeito de dominância comum apresenta dimensão equivalente ao número de famílias de irmãos germanos e a variância deste vetor contempla  $\frac{1}{4}$  da variância de dominância, ignorando a epistasia. Assim, a variância devida ao efeito de dominância comum deve ser multiplicada por quatro para obtenção da variância de dominância. Este procedimento, entretanto, não permite a predição dos efeitos de dominância para cada indivíduo, mas, prediz os efeitos aditivos livres das influências dos efeitos de dominância. O arquivo de dados deve estar organizado conforme especificado a seguir:

O modelo linear misto passa a ser dado por  $y = Xb + Za + Wc + Sf + e$ , em que f é o efeito de dominância da família de irmãos germanos e S a matriz de incidência deste efeito. A variância de f contempla  $(1/4) \sigma_d^2$ .

Os subprogramas a serem utilizados pelo DFREML são o DFPREP e o DFUNI.

#### 4.2.2 Avaliação de progênies de irmãos germanos obtidas sob cruzamentos dialélicos, fatoriais ou hierárquicos, no delineamento em blocos ao acaso, com várias plantas por parcela, várias medições por indivíduo, um só caráter e uma só população

**Modelo linear misto (modelo aditivo-dominante univariado, de repetibilidade)**

$$y = Xb + Za + Wc + Sf + Tp + e$$

#### Estruturas de médias e variâncias

$$E \begin{bmatrix} y \\ a \\ c \\ f \\ p \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} y \\ a \\ c \\ f \\ p \\ e \end{bmatrix} = \begin{bmatrix} V & ZA\sigma_a^2 & WI\sigma_c^2 & SI\sigma_f^2 & TI\sigma_p^2 & I\sigma_e^2 \\ A\sigma_a^2 Z' & A\sigma_a^2 & 0 & 0 & 0 & 0 \\ I\sigma_c^2 W' & 0 & I\sigma_c^2 & 0 & 0 & 0 \\ I\sigma_f^2 S' & 0 & 0 & I\sigma_f^2 & 0 & 0 \\ I\sigma_p^2 T' & 0 & 0 & 0 & I\sigma_p^2 & 0 \\ I\sigma_e^2 & 0 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

## Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'S & X'T \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'S & Z'T \\ W'X & W'Z & W'W + I\lambda_2 & W'S & W'T \\ S'X & S'Z & S'W & S'S + I\lambda_3 & S'T \\ T'X & T'Z & T'W & T'S & T'T + I\lambda_4 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{f} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ S'y \\ T'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-\rho}{h^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1-\rho}{c^2}; \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_f^2} = \frac{1-\rho}{0,25(h_a^2-h^2)}; \quad \lambda_4 = \frac{\sigma_e^2}{\sigma_p^2} = \frac{1-\rho}{p^2}$$

**Estimadores iterativos de componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}' Z'y - \hat{c}' W'y - \hat{f}' S'y - \hat{p}' T'y]/[N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \operatorname{tr}(A^{-1} C^{22})]/q$$

$$\hat{\sigma}_c^2 = [\hat{c}' c + \hat{\sigma}_e^2 \operatorname{tr} C^{33}]/s$$

$$\hat{\sigma}_f^2 = [\hat{f}' f + \hat{\sigma}_e^2 \operatorname{tr} C^{44}]/v$$

$\hat{\sigma}_p^2 = [\hat{p}' \hat{p} + \hat{\sigma}_e^2 \operatorname{tr} C^{55}]/s$ , em que v é o número de famílias de irmãos germanos.

**Seqüência de colunas no arquivo de dados para análise no software DFREML**

Indivíduo Pai Mãe Bloco-Medição Parcela Família Permanente Variável 1 ... Variável n

Devem ser utilizados os subprogramas DFPREP e DFUNI.

**4.2.3 Avaliação de progêneres de irmãos germanos obtidas sob cruzamentos dialélicos, fatoriais ou hierárquicos, no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo e um só caráter, avaliado em vários experimentos, com tratamentos comuns**

**Modelo linear misto (modelo aditivo-dominante multivariado)**

Considerando o caso bivariado, tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

**Distribuições e estruturas de médias e variâncias**

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ f_1 \\ f_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ f_1 \\ f_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{f_1}^2 & I\sigma_{f_{12}} & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{f_{12}} & I\sigma_{f_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}, \text{ em}$$

que:

$\sigma_{f_{12}}$ : covariância entre os efeitos de dominância associado às famílias de irmãos

germanos nos ambientes 1 e 2 (equivalente a  $\frac{1}{4} \sigma_d^2$ , livre da interação).

## Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}S \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}S \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}S \\ S'R^{-1}X & S'R^{-1}Z & S'R^{-1}W & S'R^{-1}S + F^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{f} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ S'R^{-1}y \end{bmatrix}, \text{ em}$$

que:

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}; \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix}; \quad \hat{f} = \begin{bmatrix} \hat{f}_1 \\ \hat{f}_2 \end{bmatrix};$$

$$R^{-1} = R_O^{-1} \otimes I; \quad G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C^{-1} = C_O^{-1} \otimes I; \quad F^{-1} = F_O^{-1} \otimes I$$

$$R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad F_O = \begin{bmatrix} \sigma_{f_1}^2 & \sigma_{f_{12}} \\ \sigma_{f_{12}} & \sigma_{f_2}^2 \end{bmatrix}$$

## Estimadores dos componentes de variância por REML

São recomendados os algoritmos AI (preferencialmente este) e DF. A função geral a ser maximizada é aquela descrita no item 2.

## Seqüência de colunas no arquivo de dados para análise no software DFREML

Experimento Indivíduo Pai Mãe Bloco Parcela Família Variável 1 . . . Variável n

São necessários dois arquivos, o de dados, conforme estrutura apresentada acima e um de pedigree formado pelas colunas **Indivíduo** **Pai** **Mãe**. Devem ser executados os subprogramas DFPREP e DXMUX.

**4.2.4 Avaliação de progêneres de irmãos germanos obtidas sob cruzamentos dialélicos, fatoriais ou hierárquicos, no delineamento em blocos ao acaso, com várias plantas por parcela, várias medições por indivíduo, um só caráter, uma só população, avaliada em vários experimentos com progêneres ou tratamentos comuns**

**Modelo linear misto (modelo aditivo-dominante multivariado, de repetibilidade)**

$y = Xb + Za + Wc + Sf + Tp + e$ , que para o caso bivariado equivale a:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} + \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

**Estruturas de médias e variâncias**

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ f_1 \\ f_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ f_1 \\ f_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_1 a_2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_1 a_2} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{f_1}^2 & I\sigma_{f_1 f_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{f_1 f_2} & I\sigma_{f_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{p_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{p_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

**Equações de modelo misto**

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}S & X'R^{-1}T \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}S & Z'R^{-1}T \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}S & W'R^{-1}T \\ S'R^{-1}X & S'R^{-1}Z & S'R^{-1}W & S'R^{-1}S + F^{-1} & S'R^{-1}T \\ T'R^{-1}X & T'R^{-1}Z & T'R^{-1}W & T'R^{-1}S & T'R^{-1}T + P^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{f} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ S'R^{-1}y \\ T'R^{-1}y \end{bmatrix}$$

em que:

$$R^{-1} = R_O^{-1} \otimes I; \quad G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C^{-1} = C_O^{-1} \otimes I; \quad F^{-1} = F_O^{-1} \otimes I; \quad P^{-1} = P_O^{-1} \otimes I$$

$$R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad F_O = \begin{bmatrix} \sigma_{f_1}^2 & \sigma_{f_{12}} \\ \sigma_{f_{12}} & \sigma_{f_2}^2 \end{bmatrix}; \quad P_O = \begin{bmatrix} \sigma_{p_1}^2 & 0 \\ 0 & \sigma_{p_2}^2 \end{bmatrix}$$

### **Estimadores dos componentes de variância por REML**

Neste caso, é recomendado o algoritmo AI. A função geral a ser maximizada é aquela descrita no item 2.

### **Software**

Para esta situação recomenda-se o uso do software ASREML, o qual exige um arquivo de programa específico com extensão. as.

## **4.3 Delineamento em blocos ao acaso, progênieis de polinização aberta, várias populações**

### **4.3.1 Avaliação de progênieis de meios irmãos (polinização aberta) de várias populações (procedências), no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo e um só caráter**

Nesta situação, duas formas de experimentação existem: (i) progênieis alocadas hierarquicamente dentro de procedências (arranjo de famílias compactas); (ii) progênieis de várias procedências alocadas aleatoriamente dentro de cada bloco.

Adicionalmente, existem duas opções para se tratar o efeito de populações: (a) efeito fixo; (b) efeito aleatório. Na situação (ii) (a), ou seja, famílias aleatorizadas dentro de blocos e procedência como efeito fixo, basta ajustar no vetor de efeitos fixos, os efeitos de blocos e de procedências, sendo que a interação de efeito fixo, procedência x bloco, tende a ser desprezível. Procedendo desta forma, as observações individuais são ajustadas para estimativas BLUE dos efeitos de bloco e procedências e basta adotar o modelo, estimadores e preditores descritos no tópico 4.1.1.

Na situação (i) (a), pode-se, vantajosamente, no procedimento BLUP, ajustar as observações individuais para as estimativas BLUE dos efeitos das procedências nos blocos, ou seja, considerar o ambiente homogêneo da grande parcela de procedência como o efeito fixo para o qual as observações de campo devam ser ajustadas. Neste caso, deve-se ajustar a combinação bloco-procedência como efeito fixo (com  $b p$  níveis, em que  $b$  é o número de blocos e

p o número de procedências) e também empregar o modelo, estimadores e preditores descritos no tópico 4.1.1.

Quando se consideram as procedências tanto como efeitos genéticos fixos ou aleatórios, deve-se, para efeito da seleção de indivíduos, somar os valores genéticos individuais preditos aos efeitos de procedências. Procedendo desta forma, os indivíduos de diferentes procedências podem ser comparados diretamente por seus novos valores genéticos preditos.

Na situação (i) (b) tem-se os modelos, estimadores e preditores descritos a seguir.

### **Modelo linear misto (modelo aditivo univariado, multi-populações)**

$y = Xb + Za + Wc_1 + Qr + Uc_2 + e$ , em que:

$y, b, a, c_1, r, c_2$  e  $e$ : vetores de dados, dos efeitos fixos (blocos), dos efeitos aleatórios genéticos aditivos, dos efeitos aleatórios de parcelas referentes a progênies (subparcelas), de populações (ou raça), dos efeitos aleatórios de parcelas referentes a procedências e erros aleatórios, respectivamente.

$X, Z, W, Q$  e  $U$ : matrizes de incidência para  $b, a, c_1, r, c_2$  e  $e$ , respectivamente.

### **Distribuições e estruturas de médias e variâncias**

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

$$a \mid A, \sigma_a^2 \sim N(0, A \sigma_a^2)$$

$$c_1 \mid \sigma_{c_1}^2 \sim N(0, I \sigma_{c_1}^2)$$

$$c_2 \mid \sigma_{c_2}^2 \sim N(0, I \sigma_{c_2}^2)$$

$$r \mid \sigma_r^2 \sim N(0, I \sigma_r^2)$$

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

$$\text{Cov}(a, c'_1) = 0; \quad \text{Cov}(a, r') = 0; \quad \text{Cov}(a, c'_2) = 0; \quad \text{Cov}(a, e') = 0$$

$$\text{Cov}(c_1, r') = 0; \quad \text{Cov}(c_1, c'_2) = 0; \quad \text{Cov}(c_1, e') = 0; \quad \text{Cov}(r, c'_2) = 0$$

$$\text{Cov}(r, e') = 0; \quad \text{Cov}(c_2, e') = 0, \text{ ou seja:}$$

$$E \begin{bmatrix} y \\ a \\ c_1 \\ r \\ c_2 \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad e \quad Var \begin{bmatrix} a \\ c_1 \\ r \\ c_2 \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 & 0 & 0 & 0 \\ 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_r^2 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

$$Var(y) = ZA\sigma_a^2Z' + WI\sigma_{c_1}^2W' + QI\sigma_r^2Q' + UI\sigma_{c_2}^2U' + I\sigma_e^2.$$

### Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'Q & X'U \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'Q & Z'U \\ W'X & W'Z & W'W + I\lambda_2 & W'Q & W'U \\ Q'X & Q'Z & Q'W & Q'Q + I\lambda_3 & Q'U \\ U'X & U'Z & U'W & U'Q & U'U + I\lambda_4 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{r} \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ Q'y \\ U'y \end{bmatrix}, \quad \text{em}$$

que:

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2}{h^2};$$

$$\lambda_2 = \frac{\sigma_e^2}{\sigma_{c_1}^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2}{c_1^2};$$

$$\lambda_3 = \frac{\sigma_e^2}{\sigma_r^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2}{r^2};$$

$$\lambda_4 = \frac{\sigma_e^2}{\sigma_{c_2}^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2}{c^2}.$$

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_e^2}$ : herdabilidade individual no sentido restrito;

$c_1^2 = \frac{\sigma_{c_1}^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_e^2}$ : correlação entre indivíduos devido ao ambiente comum da parcela referente a progêneres (subparcelas);

$$r^2 = \frac{\sigma_r^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_e^2}$$

: correlação fenotípica intraclasse entre indivíduos de uma mesma procedência, em diferentes blocos;

$$c_2^2 = \frac{\sigma_{c_2}^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_e^2}$$

: correlação devida ao ambiente comum da parcela de procedência.

### **Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y' y - \hat{b}' X' y - \hat{a}' Z' y - \hat{c}'_1 W' y - \hat{r}' Q' y - \hat{c}'_2 U' y] / [N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \operatorname{tr} (A^{-1} C^{22})] / q$$

$$\hat{\sigma}_{c_1}^2 = [\hat{c}'_1 \hat{c}_1 + \hat{\sigma}_e^2 \operatorname{tr} C^{33}] / s_1$$

$$\hat{\sigma}_r^2 = [\hat{r}' \hat{r} + \hat{\sigma}_e^2 \operatorname{tr} C^{44}] / t$$

$$\hat{\sigma}_{c_2}^2 = [\hat{c}'_2 \hat{c}_2 + \hat{\sigma}_e^2 \operatorname{tr} C^{55}] / s_2, \text{ em que:}$$

$s_1$ ,  $t$  e  $s_2$  = número de parcelas referentes a progênies, número de procedências e número de parcelas referentes a procedência, respectivamente.

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

Indivíduo	Pai	Mãe	Bloco	Parcela <sub>1</sub>	Procedência	Parcela <sub>2</sub>	Variável 1	...	Variável n
-----------	-----	-----	-------	----------------------	-------------	----------------------	------------	-----	------------

Neste caso, é necessário apenas um arquivo o qual funciona ao mesmo tempo como arquivo de pedigree e como arquivo de dados. Deve ser executado o subprograma DFPREP e em seguida o DFUNI.

A significância da inclusão do efeito ( $c_2$ ) da parcela de procedência pode ser avaliada através do teste da razão de verossimilhança. Se este efeito não for significativo, deve-se retirá-lo do modelo e dos preditores e estimadores apresentados anteriormente. Estas mesmas considerações são válidas para a situação (ii) (b), em que a interação bloco x procedência, de efeito aleatório tende a ser não significativa, podendo ser retirada do modelo.

É importante relatar que os modelos apresentados estimam uma herdabilidade média dentro das populações. Para se estudar comparativamente as herdabilidades e coeficientes de variação genética das várias populações deve-se ajustar um modelo do tipo apresentado em 4.1.1. para cada população.

#### **4.3.2 Avaliação de progêneres de meios irmãos (polinização aberta) de várias populações (procedências), no delineamento em blocos ao acaso, com várias plantas por parcela, várias medições por indivíduo e um só caráter**

Considerando-se as situações (i) (a), (i) (b), (ii) (a) e (ii) (b) descritas no tópico 4.3.1, tem-se que:

- na situação (ii) (a), basta ajustar no vetor de efeitos fixos, os efeitos de procedência e da combinação bloco-medição e adotar o modelo, estimadores e preditores apresentados no tópico 4.1.2.
- na situação (i) (a), deve-se ajustar no vetor de efeitos fixos, o efeito da combinação bloco-procedência-medição (com bpm níveis, em que p, b e m são os números de procedências, blocos e medições, respectivamente). Neste caso, as observações individuais em uma dada medição são ajustadas para a média da procedência no bloco, em uma dada medição. O modelo, estimadores e preditores apresentados no tópico 4.1.2. devem ser empregados.

Na situação (i) (b) tem-se os modelos, estimadores e preditores descritos a seguir:

##### **Modelo linear misto (modelo aditivo univariado, multi-populações, de repetibilidade)**

$$y = Xb + Za + Wc_1 + Tp + Qr + Uc_2 + e, \text{ em que:}$$

- b : vetor de efeitos fixos (no caso, combinações bloco-medição, significando que os valores individuais em uma dada medição serão ajustados para a média do bloco na medição);
- p : vetor de efeitos permanentes (ambiente permanente dentro de parcela + efeitos genéticos não aditivos);
- T : matriz de incidência para p.

## Distribuições e estruturas de médias e variâncias

$$\begin{aligned}
 y &\mid b, V \sim N(Xb, V) \\
 a &\mid A, \sigma_a^2 \sim N(0, A\sigma_a^2) \\
 c_1 &\mid \sigma_{c_1}^2 \sim N(0, I\sigma_{c_1}^2) \\
 c_2 &\mid \sigma_{c_2}^2 \sim N(0, I\sigma_{c_2}^2) \\
 r &\mid \sigma_r^2 \sim N(0, I\sigma_r^2) \\
 p &\mid \sigma_p^2 \sim N(0, I\sigma_p^2) \\
 e &\mid \sigma_e^2 \sim N(0, I\sigma_e^2)
 \end{aligned}$$

As covariâncias entre todos os efeitos aleatórios do modelo são assumidas como nulas.

Assim:

$$E \begin{bmatrix} y \\ a \\ c_1 \\ p \\ r \\ c_2 \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad Var \begin{bmatrix} a \\ c_1 \\ p \\ r \\ c_2 \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_p^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_r^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

$$Var(y) = ZA\sigma_a^2Z' + WI\sigma_{c_1}^2W' + TI\sigma_p^2T' + QI\sigma_r^2Q' + UI\sigma_{c_2}^2U' + I\sigma_e^2.$$

## Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q & X'U \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'T & Z'Q & Z'U \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'Q & W'U \\ T'X & T'Z & T'W & T'T + I\lambda_3 & T'Q & T'U \\ Q'X & Q'Z & Q'W & Q'T & Q'Q + I\lambda_4 & Q'U \\ U'X & U'Z & U'W & U'T & U'Q & U'U + I\lambda_5 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{p} \\ \hat{r} \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ Q'y \\ U'y \end{bmatrix},$$

em que:

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - p^2}{h^2};$$

$$\lambda_2 = \frac{\sigma_e^2}{\sigma_{c_1}^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - p^2}{c_1^2};$$

$$\lambda_3 = \frac{\sigma_e^2}{\sigma_p^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - p^2}{p^2};$$

$$\lambda_4 = \frac{\sigma_e^2}{\sigma_r^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - p^2}{r^2};$$

$$\lambda_5 = \frac{\sigma_e^2}{\sigma_{c_2}^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - p^2}{c^2}.$$

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_p^2 + \sigma_e^2};$$

$$c_1^2 = \frac{\sigma_{c_1}^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_p^2 + \sigma_e^2};$$

$$r^2 = \frac{\sigma_r^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_p^2 + \sigma_e^2};$$

$$c_2^2 = \frac{\sigma_{c_2}^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_p^2 + \sigma_e^2}.$$

$p^2 = \frac{\sigma_p^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_p^2 + \sigma_e^2}$ : correlação entre medidas repetidas devidas aos efeitos permanentes (ambiental + genético não aditivo).

Neste modelo, a repetibilidade é dada por  $\rho = \frac{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_p^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_p^2 + \sigma_e^2}$ .

### **Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y' y - \hat{b}' X' y - \hat{a}' Z' y - \hat{c}'_1 W' y - \hat{p}' T' y - \hat{r}' Q' y - \hat{c}'_2 U' y] / [N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \text{tr } (A^{-1} C^{22})] / q$$

$$\hat{\sigma}_{c_1}^2 = [\hat{c}'_1 \hat{c}_1 + \hat{\sigma}_e^2 \text{tr } C^{33}] / s_1;$$

$$\hat{\sigma}_p^2 = [\hat{p}' \hat{p} + \hat{\sigma}_e^2 \text{tr } C^{44}] / q$$

$$\hat{\sigma}_r^2 = [\hat{r}' \hat{r} + \hat{\sigma}_e^2 \text{tr } C^{55}] / t;$$

$$\hat{\sigma}_{c_2}^2 = [\hat{c}'_2 \hat{c}_2 + \hat{\sigma}_e^2 \text{tr } C^{66}] / s_2.$$

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

Indivíduo Pai Mãe Bloco Parcела<sub>1</sub> Permanente Procedência Parcела<sub>2</sub> Variável 1 . . . Variável n

É necessário apenas um arquivo, o qual funciona ao mesmo tempo como arquivo de pedigree e de dados. Deve ser executado o subprogramma DFPREP e, em seguida, o DFUNI.

Na situação (ii) (b), a interação aleatória procedência x bloco tende a ser não significativa, sendo que o modelo, preditores e estimadores apresentados para o caso (i) (b) podem ser utilizados, porém eliminando-se o componente  $c_2$  e suas variâncias. A significância ou não de  $c_2$  pode ser verificada pelo teste da razão de verossimilhança.

Tanto na situação (i) (b) quanto (ii) (b), as interações procedência x medição, progênie/procedência x medição, procedência x medição x bloco e progênie/procedência x medição x bloco são incluídas no vetor residual  $e$ , o qual contempla o efeito total de ambiente temporário. É importante relatar também que o próprio modelo de repetibilidade assume que a interação com medições seja aproximadamente nula (ou seja, assume correlação genética igual tendendo a 1, através das medições).

Em todas as situações apresentadas neste tópico, o mérito genético total de cada indivíduo é dado pelo somatório do efeito genético aditivo predito com o efeito da procedência a que pertence tal indivíduo.

#### **4.3.3 Avaliação de progêneres de meios irmãos (polinização aberta) de várias populações (procedências), no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo e um só caráter, avaliado em vários experimentos com algumas progêneres ou tratamentos comuns**

Para as situações descritas em (i) (a) e (ii) (a) do tópico 4.3.1, basta adotar o modelo, estimadores e preditores descritos no tópico 4.1.3. Para as situações (i)(b) ou (ii) (b) incluindo a interação procedência x bloco (efeito  $c_2$ ), tem-se os modelos, estimadores e preditores descritos a seguir.

##### **Modelo linear misto (modelo aditivo bivariado, multi-populações)**

$y = Xb + Za + Wc_1 + Qr + Uc_2 + e$ , que para o caso bivariado equivale a:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \end{bmatrix} + \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} + \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} c_{21} \\ c_{22} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$y, b, a, c_1, r, c_2$  e  $e$ : vetores de dados, dos efeitos fixos (blocos), dos efeitos aleatórios genéticos aditivos, dos efeitos aleatórios de parcelas referentes a progêneres, dos efeitos aleatórios de populações, dos efeitos aleatórios de parcelas referentes a procedências e de erros aleatórios, respectivamente, os quais são desdobrados para os experimentos (locais) 1 e 2.

$X, Z, W, Q$  e  $U$ : matrizes de incidência para  $b, a, c_1, r, c_2$  e  $e$ , respectivamente.

## Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_{11} \\ c_{12} \\ r_1 \\ r_2 \\ c_{21} \\ c_{22} \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_{11} \\ c_{12} \\ r_1 \\ r_2 \\ c_{21} \\ c_{22} \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_1 a_2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_1 a_2} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_{11}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_{12}}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{r_1}^2 & I\sigma_{r_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{r_1} & I\sigma_{r_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{c_{21}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{c_{22}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

$$Var(y_1) = V_1 = Z_1 A \sigma_{a_1}^2 Z'_1 + W_1 I \sigma_{c_{11}}^2 W'_1 + Q_1 I \sigma_{r_1}^2 Q'_1 + U_1 I \sigma_{c_{21}}^2 U'_1 + I \sigma_{e_1}^2$$

$$Var(y_2) = V_2 = Z_2 A \sigma_{a_2}^2 Z'_2 + W_2 I \sigma_{c_{12}}^2 W'_2 + Q_2 I \sigma_{r_2}^2 Q'_2 + U_2 I \sigma_{c_{22}}^2 U'_2 + I \sigma_{e_2}^2$$

em que:

$\sigma_{a_1}^2$  e  $\sigma_{a_2}^2$ : variâncias genéticas aditivas nos locais 1 e 2, respectivamente;

$\sigma_{a_{12}}$ : covariância genética aditiva entre os locais 1 e 2 ou variância aditiva livre da interação genótipo x ambiente;

$\sigma_{c_{11}}^2$  e  $\sigma_{c_{12}}^2$ : variâncias entre parcelas (referentes a progênies) nos locais 1 e 2, respectivamente;

$\sigma_{r_1}^2$  e  $\sigma_{r_2}^2$ : variâncias entre procedências nos locais 1 e 2, respectivamente;

$\sigma_{r_{12}}$ : covariância genética entre os locais 1 e 2, ao nível do efeito de procedências;

$\sigma_{c_{21}}^2$  e  $\sigma_{c_{22}}^2$ : variâncias entre parcelas (referentes a procedências) nos locais 1 e 2, respectivamente;

$\sigma_{e_1}^2$  e  $\sigma_{e_2}^2$ : variâncias residuais nos locais 1 e 2, respectivamente;

## Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}Q & X'R^{-1}U \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}Q & Z'R^{-1}U \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C_1^{-1} & W'R^{-1}Q & W'R^{-1}U \\ Q'R^{-1}X & Q'R^{-1}Z & Q'R^{-1}W & Q'R^{-1}Q + \mathfrak{R}^{-1} & Q'R^{-1}U \\ U'R^{-1}X & U'R^{-1}Z & U'R^{-1}W & U'R^{-1}Q & U'R^{-1}U + C_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{r} \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ Q'R^{-1}y \\ U'R^{-1}y \end{bmatrix},$$

em que:

$$G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C_1^{-1} = C_{10}^{-1} \otimes I; \quad \mathfrak{R}^{-1} = \mathfrak{R}_O^{-1} \otimes I; \quad C_2^{-1} = C_{20}^{-1} \otimes I; \quad R^{-1} = R_O^{-1} \otimes I$$

$$G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad \mathfrak{R}_O = \begin{bmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_{12}} \\ \sigma_{\eta_{12}} & \sigma_{\eta_2}^2 \end{bmatrix}; \quad C_{10} = \begin{bmatrix} \sigma_{c_{11}}^2 & 0 \\ 0 & \sigma_{c_{12}}^2 \end{bmatrix}; \quad C_{20} = \begin{bmatrix} \sigma_{c_{21}}^2 & 0 \\ 0 & \sigma_{c_{22}}^2 \end{bmatrix}; \quad R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}$$

Os componentes de variância estão associados aos parâmetros

$h^2$ ,  $c_1^2$ ,  $c_2^2$ ,  $\rho_{a_{ij}}$  e  $\rho_{r_{ij}}$ , da seguinte maneira:

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{c_{1i}}^2 = c_{1i}^2 \sigma_{y_i}^2; \quad \sigma_{c_{2i}}^2 = c_{2i}^2 \sigma_{y_i}^2; \quad \sigma_{e_i}^2 = (1 - h_i^2 - c_{1i}^2 - c_{2i}^2 - r_i^2) \sigma_{y_i}^2;$$

$$\sigma_{r_i}^2 = r_i^2 \sigma_{y_i}^2; \quad \sigma_{a_{ij}} = \rho_{a_{ij}} \sigma_{a_i} \sigma_{a_j}; \quad \sigma_{r_{ij}} = \rho_{r_{ij}} \sigma_{r_i} \sigma_{r_j}, \text{ em que:}$$

$\rho_{a_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i} \sigma_{a_j}}$ : correlação genética entre o desempenho dos indivíduos nos locais i e j;

$\rho_{r_{ij}} = \frac{\sigma_{r_{ij}}}{\sigma_{r_i} \sigma_{r_j}}$ : correlação genética entre o desempenho das procedências nos locais i e j;

$\sigma_{y_i}^2$ : variância fenotípica ao nível de indivíduo no local i.

A variância da interação indivíduo x local, para o caso balanceado, é dada por:

$$\sigma_{ae}^2 = \frac{1}{2}(\sigma_{a_i} - \sigma_{a_j})^2 + (1 - \rho_{a_{ij}}) \sigma_{a_i} \sigma_{a_j}.$$

A variância da interação procedência x local, para o caso balanceado, é dada por:

$$\sigma_{re}^2 = \frac{1}{2}(\sigma_{r_i} - \sigma_{r_j})^2 + (1 - \rho_{r_{ij}}) \sigma_{r_i} \sigma_{r_j}.$$

#### **Estimadores dos componentes de variância por REML**

Nesta situação, devido à complexidade do modelo e ao elevado número de componentes de variância a serem estimados, os algoritmos recomendados são o DF e o AI (preferencialmente este). Neste caso, a função geral a ser maximizada é aquela descrita no item 2.

#### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

Experimento Indivíduo Pai Mãe Bloco Parcela<sub>1</sub> Procedência Parcela<sub>2</sub> Variável 1 . . . Variável n

#### **4.3.4 Avaliação de progênies de meios irmãos (polinização aberta) de várias populações (procedências), no delineamento em blocos ao acaso, com várias plantas por parcela, várias medições por indivíduo e um só caráter, avaliado em vários experimentos com algumas progênies ou tratamentos comuns**

Para as situações descritas em (i) (a) e (ii) (a) do tópico 4.3.1, basta adotar o modelo, estimadores e preditores descritos no tópico 4.1.4. Para as situações (i)(b) ou (ii) (b) incluindo a interação procedência x bloco (efeito c<sub>2</sub>), tem-se os modelos, estimadores e preditores descritos a seguir.

## Modelo linear misto (modelo aditivo bivariado, multi-populações, de repetibilidade)

$y = Xb + Za + Wc_1 + Qr + Tp + e$ , que para o caso bivariado equivale a:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \end{bmatrix} + \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} + \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$y, b, a, c_1, r, p$  e  $e$ : vetores de dados, dos efeitos fixos (blocos), dos efeitos aleatórios genéticos aditivos, dos efeitos aleatórios de parcelas referentes a progênies, dos efeitos aleatórios de populações, dos efeitos aleatórios permanentes e de erros aleatórios, respectivamente, os quais são desdobrados para os experimentos (locais) 1 e 2.

$X, Z, W, Q$  e  $T$ : matrizes de incidência para  $b, a, c_1, r, p$  e  $e$ , respectivamente.

### Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ X_1 b_1 \\ X_2 b_2 \\ a_1 \\ a_2 \\ c_{11} \\ c_{12} \\ r_1 \\ r_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_{11} \\ c_{12} \\ r_1 \\ r_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_1 a_2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_1 a_2} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_{11}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_{12}}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_r^2 & I\sigma_{r_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{r_1} & I\sigma_{r_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{p_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{p_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

$$Var(y_1) = V_1 = Z_1 A \sigma_{a_1}^2 Z'_1 + W_1 I \sigma_{c_{11}}^2 W'_1 + Q_1 I \sigma_{r_1}^2 Q'_1 + T_1 I \sigma_{p_1}^2 T'_1 + I \sigma_{e_1}^2$$

$$Var(y_2) = V_2 = Z_2 A \sigma_{a_2}^2 Z'_2 + W_2 I \sigma_{c_{12}}^2 W'_2 + Q_2 I \sigma_{r_2}^2 Q'_2 + T_2 I \sigma_{p_2}^2 T'_2 + I \sigma_{e_2}^2, \text{ em que:}$$

$\sigma_{a_1}^2$  e  $\sigma_{a_2}^2$ : variâncias genéticas aditivas nos locais 1 e 2, respectivamente;

$\sigma_{a_1 a_2}$ : covariância genética aditiva entre os locais 1 e 2 ou variância aditiva livre da interação;

$\sigma_{c_{11}}^2$  e  $\sigma_{c_{12}}^2$ : variâncias entre parcelas (referentes a progênies) nos locais 1 e 2, respectivamente;

$\sigma_{\eta_1}^2$  e  $\sigma_{\eta_2}^2$ : variâncias entre procedências nos locais 1 e 2, respectivamente;

$\sigma_{\eta_{12}}$ : covariância genética entre os locais 1 e 2, ao nível do efeito de procedências;

$\sigma_{p_1}^2$  e  $\sigma_{p_2}^2$ : variâncias permanentes nos locais 1 e 2, respectivamente;

$\sigma_{e_1}^2$  e  $\sigma_{e_2}^2$ : variâncias residuais nos locais 1 e 2, respectivamente;

### Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}Q & X'R^{-1}T \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}Q & Z'R^{-1}T \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C_1^{-1} & W'R^{-1}Q & W'R^{-1}T \\ Q'R^{-1}X & Q'R^{-1}Z & Q'R^{-1}W & Q'R^{-1}Q + \mathfrak{R}^{-1} & Q'R^{-1}T \\ T'R^{-1}X & T'R^{-1}Z & T'R^{-1}W & T'R^{-1}Q & T'R^{-1}T + C_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{r} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ Q'R^{-1}y \\ T'R^{-1}y \end{bmatrix},$$

em que:

$$G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C_1^{-1} = C_{10}^{-1} \otimes I; \quad \mathfrak{R}^{-1} = \mathfrak{R}_O^{-1} \otimes I; \quad P^{-1} = P_0^{-1} \otimes I; \quad R^{-1} = R_O^{-1} \otimes I$$

$$G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad \mathfrak{R}_O = \begin{bmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_{12}} \\ \sigma_{\eta_{12}} & \sigma_{\eta_2}^2 \end{bmatrix}; \quad C_{10} = \begin{bmatrix} \sigma_{c_{11}}^2 & 0 \\ 0 & \sigma_{c_{12}}^2 \end{bmatrix}; \quad P_0 = \begin{bmatrix} \sigma_{p_1}^2 & 0 \\ 0 & \sigma_{p_2}^2 \end{bmatrix}; \quad R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}$$

Os componentes de variância estão associados aos parâmetros  $h^2$ ,  $c_1^2$ ,  $p^2$ ,  $\rho_{a_{ij}}$  e  $\rho_{r_{ij}}$ , da seguinte maneira:

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{c_{1i}}^2 = c_{1i}^2 \sigma_{y_i}^2; \quad \sigma_{p_i}^2 = p_i^2 \sigma_{y_i}^2; \quad \sigma_{e_i}^2 = (1 - h_i^2 - c_{1i}^2 - p_i^2) \sigma_{y_i}^2;$$

$$\sigma_{r_i}^2 = r_i^2 \sigma_{y_i}^2; \quad \sigma_{a_{ij}} = \rho_{a_{ij}} \sigma_{a_i} \sigma_{a_j}; \quad \sigma_{r_{ij}} = \rho_{r_{ij}} \sigma_{r_i} \sigma_{r_j}, \text{ em que:}$$

$\rho_{a_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i} \sigma_{a_j}}$ : correlação genética entre o desempenho dos indivíduos nos locais i e j;

$\rho_{r_{ij}} = \frac{\sigma_{r_{ij}}}{\sigma_{r_i} \sigma_{r_j}}$ : correlação genética entre o desempenho das procedências nos locais i e j;

$\sigma_{y_i}^2$ : variância fenotípica individual no local i.

A variância da interação indivíduo x local, para o caso balanceado, é dada por:

$$\sigma_{ae}^2 = \frac{1}{2}(\sigma_{a_i} - \sigma_{a_j})^2 + (1 - \rho_{a_{ij}}) \sigma_{a_i} \sigma_{a_j}.$$

A variância da interação procedência x local, para o caso balanceado, é dada por:

$$\sigma_{re}^2 = \frac{1}{2}(\sigma_{r_i} - \sigma_{r_j})^2 + (1 - \rho_{r_{ij}}) \sigma_{r_i} \sigma_{r_j}.$$

### **Estimadores dos componentes de variância por REML**

Nesta situação, devido à complexidade do modelo e ao elevado número de componentes de variância a serem estimados, os algoritmos recomendados são o DF e o AI (preferencialmente este). Neste caso, a função geral a ser maximizada é aquela descrita no item 2.

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

Experimento Indivíduo Pai Mãe Bloco Parcela<sub>1</sub> Procedência Parcela<sub>2</sub> Variável 1 . . . Variável n

## 4.4 Delineamento em látice, progêneres de polinização aberta, uma população

### 4.4.1 Avaliação de progêneres de meios irmãos, no delineamento em látice, com várias plantas por parcela, uma medição por indivíduo, um só caráter e uma só população

#### Modelo linear misto (modelo aditivo univariado, em látice)

$$y = Xb + Za + Wc + H\eta + e, \text{ em que}$$

$y, b, a, c, \eta$  e  $e$ : vetores de dados, dos efeitos de repetições (fixos), de efeitos genéticos aditivos (aleatórios), dos efeitos de parcela (aleatório), dos efeitos aleatórios de blocos dentro de repetições e de erros aleatórios, respectivamente.

$X, Z, W$  e  $H$ : matrizes de incidência para  $b, a, c$  e  $\eta$ , respectivamente.

#### Distribuições e estruturas de médias e variâncias

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

$$a|A, \sigma_a^2 \sim N(0, A \sigma_a^2)$$

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

$$\partial|\sigma_\partial^2 \sim N(0, I \sigma_\partial^2)$$

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

As covariâncias entre todos os efeitos aleatórios do modelo são assumidas como nulas.

Assim:

$$E \begin{bmatrix} y \\ a \\ c \\ \partial \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad e \quad Var \begin{bmatrix} y \\ a \\ c \\ \partial \\ e \end{bmatrix} = \begin{bmatrix} V & ZG & WC & HB & R \\ GZ' & G & 0 & 0 & 0 \\ CW' & 0 & C & 0 & 0 \\ BH' & 0 & 0 & B & 0 \\ R & 0 & 0 & 0 & R \end{bmatrix}, \text{ em que:}$$

$$G = A \sigma_A^2$$

$$C = I \sigma_c^2$$

$$B = I \sigma_{\partial}^2$$

$$R = I \sigma_e^2$$

$$V = ZA \sigma_a^2 Z' + WI \sigma_c^2 W' + H'I\sigma_{\partial}^2 H + I \sigma_e^2 = ZGZ' + WCW' + HBH' + R.$$

### Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'H \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'H \\ W'X & W'Z & W'W + I\lambda_2 & W'H \\ H'X & H'Z & H'W & H'H + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{\partial} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ H'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2 - c^2 - \partial^2}{h^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1-h^2 - c^2 - \partial^2}{c^2} \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_{\partial}^2} = \frac{1-h^2 - c^2 - \partial^2}{\partial^2}$$

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2 + \sigma_{\partial}^2}$ : herdabilidade individual no sentido restrito, na repetição;

$c^2 = \sigma_c^2 / (\sigma_a^2 + \sigma_c^2 + \sigma_e^2 + \sigma_{\partial}^2)$ : correlação devida ao ambiente comum da parcela;

$\partial^2 = \sigma_{\partial}^2 / (\sigma_a^2 + \sigma_c^2 + \sigma_e^2 + \sigma_{\partial}^2)$ : correlação devida ao ambiente comum do bloco dentro de repetição;

$\sigma_a^2$ : variância genética aditiva;

$\sigma_c^2$ : variância entre parcelas;

$\sigma_{\partial}^2$ : variância entre blocos dentro de repetições;

$\sigma_e^2$ : variância residual (ambiental dentro de parcelas + não aditiva);

A: matriz de correlação genética aditiva entre os indivíduos em avaliação.

## **Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}' Z'y - \hat{c}' W'y - \hat{\delta}' H'y] / [N - r(x)];$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \operatorname{tr}(A^{-1} C^{22})] / q;$$

$$\hat{\sigma}_c^2 = [\hat{c}' c + \hat{\sigma}_e^2 \operatorname{tr} C^{33}] / s;$$

$$\hat{\sigma}_{\delta}^2 = [\hat{\delta}' \hat{\delta} + \hat{\sigma}_e^2 \operatorname{tr} C^{44}] / \eta, \text{ em que:}$$

$C^{22}$ ,  $C^{33}$  e  $C^{44}$  advém da inversa da matriz dos coeficientes das equações de modelo misto.

$\operatorname{tr}$  : operador traço matricial;

$r(x)$  : posto da matriz  $X$ ;

$N$ ,  $q$ ,  $s$ ,  $h$  : número total de dados, número de indivíduos, número de parcelas e número de blocos, respectivamente.

## **Seqüência de colunas no arquivo de dados para análise no software DFREML**

Indivíduo    Pai    Mãe    Repetição    Parcela    Bloco    Variável 1    . . .    Variável n

Neste caso, é necessário apenas um arquivo o qual funciona ao mesmo tempo como arquivo de pedigree e como arquivo de dados. Deve ser executado o subprograma DFPREP e, em seguida, o DFUNI.

### **4.4.2 Avaliação de progêneres de meios irmãos, no delineamento em látice, com várias plantas por parcela, várias medições por indivíduo, um só caráter e uma só população**

#### **Modelo linear misto (modelo aditivo univariado, de repetibilidade, em látice)**

$y = Xb + Za + Wc + Tp + H\pi + e$ , em que:

$p$ : vetor de efeitos permanentes (ambiente permanente dentro de parcela + efeitos genéticos não aditivos);

$T$ : matriz de incidência para  $p$ .

Neste modelo, pode ser ajustado um único efeito (denominado combinação repetição-medição), procedimento este que é estatisticamente correto e computacionalmente desejável e necessário.

### Distribuições e estruturas de médias e variâncias

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

$$a|A, \sigma_a^2 \sim N(0, \sigma_a^2)$$

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

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

$$\partial|\sigma_\partial^2 \sim N(0, I \sigma_\partial^2)$$

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

As covariâncias entre todos os efeitos aleatórios dos modelos são assumidas como nulas.

Assim:

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

$$P = I \sigma_p^2$$

$$V = ZA\sigma_a^2Z' + WI\sigma_c^2W' + TI\sigma_p^2T' + HI\sigma_\partial^2H' + I\sigma_e^2.$$

## Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'T & X'H \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'T & Z'H \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'H \\ T'X & T'Z & T'W & T'T + I\lambda_3 & T'H \\ H'X & H'Z & H'W & H'T & H'H + I\lambda_4 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \\ \hat{\partial} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ H'y \end{bmatrix}, \quad \text{em}$$

que:

$$\lambda_1 = \frac{1-\rho}{h^2} = \frac{\sigma_e^2}{\sigma_a^2}; \quad \lambda_2 = \frac{1-\rho}{c^2} = \frac{\sigma_e^2}{\sigma_c^2}; \quad \lambda_3 = \frac{1-\rho}{p^2} = \frac{\sigma_e^2}{\sigma_p^2}; \quad \lambda_4 = \frac{1-\rho}{\partial^2} = \frac{\sigma_e^2}{\sigma_\partial^2}.$$

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2 + \sigma_\partial^2}$ : herdabilidade individual no sentido restrito, na repetição, em uma dada medição;

$\rho = \frac{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_\partial^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2 + \sigma_\partial^2}$ : repetibilidade individual na repetição;

$p^2 = \frac{\sigma_p^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2 + \sigma_\partial^2}$ : coeficiente de determinação dos efeitos permanentes dentro de parcela;

$c^2 = \frac{\sigma_c^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2 + \sigma_\partial^2}$ : correlação devida ao ambiente comum de parcela;

$\partial^2 = \frac{\sigma_\partial^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2 + \sigma_\partial^2}$ : correlação devida ao ambiente comum do bloco.

## Estimadores iterativos dos componentes de variância por REML via algoritmo EM

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}' Z'y - \hat{c}' W'y - \hat{p}' T'y - \hat{\partial}' H'y] / [N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \operatorname{tr}(A^{-1} C^{22})] / q$$

$$\hat{\sigma}_c^2 = [\hat{c}' c + \hat{\sigma}_e^2 \operatorname{tr} C^{33}] / s$$

$$\hat{\sigma}_p^2 = [\hat{p}' \hat{p} + \hat{\sigma}_e^2 \operatorname{tr} C^{44}] / q$$

$$\hat{\sigma}_{\partial}^2 = [\hat{\partial}' \hat{\partial} + \hat{\sigma}_e^2 \operatorname{tr} C^{55}] / \eta, \text{ em que:}$$

$C^{22}, C^{33}, C^{44}$  e  $C^{55}$  advém da inversa da matriz dos coeficientes das equações de modelo misto.

## Seqüência de colunas no arquivo de dados para análise no software DFREML

Indivíduo Pai Mãe Repetição-Medição Parcela Permanente Bloco Variável 1 ... Variável n

É necessário apenas um arquivo, o qual funciona simultaneamente como arquivo de dados e como arquivo de pedigree. Devem ser executados seqüencialmente os subprogramas DFPREP e DFUNI.

### 4.4.3 Avaliação de progêneres de meios irmãos, no delineamento em látice, com várias plantas por parcela, uma medição por indivíduo e um só caráter, avaliado em vários experimentos com algumas progêneres ou tratamentos comuns

#### Modelo linear misto (modelo aditivo multivariado, em látice)

Considerando o caso bivariado, tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} \begin{bmatrix} \partial_1 \\ \partial_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

## Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ \partial_1 \\ \partial_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ \partial_1 \\ \partial_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{\partial_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{\partial_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix},$$

em que:

$\sigma_{\partial_1}^2$  e  $\sigma_{\partial_2}^2$ . = variância permanente entre blocos dentro de repetições nos locais 1 e 2, respectivamente.

## Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}H \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}H \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}H \\ H'R^{-1}X & H'R^{-1}Z & H'R^{-1}W & H'R^{-1}H + B^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{\partial} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ H'R^{-1}y \end{bmatrix}, \text{ em}$$

que:

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}; \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix}; \quad \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix}; \quad \hat{\partial} = \begin{bmatrix} \hat{\partial}_1 \\ \hat{\partial}_2 \end{bmatrix}$$

$$R^{-1} = R_O^{-1} \otimes I; \quad G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C^{-1} = C_O^{-1} \otimes I; \quad B^{-1} = B_O^{-1} \otimes I$$

$$R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad B_O = \begin{bmatrix} \sigma_{\partial_1}^2 & 0 \\ 0 & \sigma_{\partial_2}^2 \end{bmatrix}$$

Os componentes de variância estão associados aos parâmetros  $h^2$ ,  $c^2$ ,  $\Psi^2$  e  $r_a$ , da seguinte maneira:

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{e_i}^2 = (1 - \rho_i) \sigma_{y_i}^2; \quad \sigma_{c_i}^2 = c_i^2 \sigma_{y_i}^2; \quad \sigma_{\partial_i}^2 = (\partial_i^2 - h_i^2 - c_i^2) \sigma_{y_i}^2; \quad \sigma_{a_{ij}} = \rho_{a_{ij}} \sigma_{a_i} \sigma_{a_j}$$

### Estimadores dos componentes de variância por REML

Nesta situação são recomendados os algoritmos AI (preferencialmente este) e DF. A função geral a ser maximizada é aquela descrita no item 2.

### Seqüência de colunas de dados para análise no software DFREML

Experimento Indivíduo Pai Mãe Repetição Parcada Bloco Variável 1 ... Variável n

São necessários dois arquivos, o de dados e um de pedigree formado pelas colunas Indivíduo Pai Mãe. Devem ser executados os subprogramas DFPREP e DXMUX.

#### 4.4.4 Avaliação de progêneros de meios irmãos, no delineamento em látice, com várias plantas por parcela, várias medições por indivíduo e um só caráter, avaliado em vários experimentos com algumas progêneros ou tratamentos comuns

**Modelo linear misto (modelo aditivo multivariado, de repetibilidade, em látice)**

$y = Xb + Za + Wc + Tp + H\Psi + e$  que, para o caso bivariado equivale a:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} \begin{bmatrix} \partial_1 \\ \partial_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

## Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ \partial_1 \\ \partial_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ \partial_1 \\ \partial_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_1} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{p_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{p_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{\partial_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{\partial_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

## Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}T & X'R^{-1}H \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}T & Z'R^{-1}H \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}T & W'R^{-1}H \\ T'R^{-1}X & T'R^{-1}Z & T'R^{-1}W & T'R^{-1}T + P^{-1} & T'R^{-1}H \\ H'R^{-1}X & H'R^{-1}Z & H'R^{-1}W & H'R^{-1}T & H'R^{-1}H + P^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \\ \hat{\partial} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ T'R^{-1}y \\ H'R^{-1}y \end{bmatrix}$$

em que:

$$R^{-1} = R_o^{-1} \otimes I; \quad G^{-1} = G_o^{-1} \otimes A^{-1}; \quad C^{-1} = C_o^{-1} \otimes I; \quad P^{-1} = P_o^{-1} \otimes I; \quad B^{-1} = B_o^{-1} \otimes I$$

$$R_o = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_o = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_2} \\ \sigma_{a_1} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_o = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad P_o = \begin{bmatrix} \sigma_{p_1}^2 & 0 \\ 0 & \sigma_{p_2}^2 \end{bmatrix}; \quad B_o = \begin{bmatrix} \sigma_{\partial_1}^2 & 0 \\ 0 & \sigma_{\partial_2}^2 \end{bmatrix}$$

## Estimadores dos componentes de variância por REML

Neste caso, é recomendado o algoritmo AI. A função geral a ser maximizada é aquela descrita no item 2.

## Software

Para esta situação recomenda-se o uso do software ASREML, o qual exige um arquivo de programa específico com extensão .as.

## **4.5. Delineamento em látice, progênie de polinização aberta, várias populações**

### **4.5.1 Avaliação de progênie de meios irmãos (polinização aberta) de várias populações (procedências), no delineamento em látice, com várias plantas por parcela, uma medição por indivíduo e um só caráter**

Para as situações descritas em (i) (a) e (ii) (a) do tópico 4.3.1, basta adotar o modelo, estimadores e preditores descritos no tópico 4.4.1. Para as situações (i)(b) ou (ii) (b) incluindo a interação procedência x bloco (efeito  $c_2$ ), tem-se os modelos, estimadores e preditores descritos a seguir.

#### **Modelo linear misto (modelo aditivo univariado, multi-populações, em látice)**

$y = Xb + Za + Wc_1 + H\P + Qr + Uc_2 + e$ , em que:

$b$  = vetor de efeitos fixos (repetições, significando que os valores individuais serão ajustados para a média da repetição);

$\P$  = vetor de efeitos aleatórios dos blocos dentro de repetições;

$H$  = matriz de incidência para  $\P$ .

#### **Distribuições e estruturas de médias e variâncias**

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

$$a \mid A, \sigma_a^2 \sim N(0, A \sigma_a^2)$$

$$c_1 \mid \sigma_{c_1}^2 \sim N(0, I \sigma_{c_1}^2)$$

$$c_2 \mid \sigma_{c_2}^2 \sim N(0, I \sigma_{c_2}^2)$$

$$r \mid \sigma_r^2 \sim N(0, I \sigma_r^2)$$

$$\partial \mid \sigma_\partial^2 \sim N(0, I \sigma_\partial^2)$$

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

As covariâncias entre todos os efeitos aleatórios do modelo são assumidas como nulas.

Assim:

$$E \begin{bmatrix} y \\ a \\ c_1 \\ \partial \\ r \\ c_2 \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad e \quad Var \begin{bmatrix} a \\ c_1 \\ \partial \\ r \\ c_2 \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{\partial}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_r^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

$$Var(y) = ZA\sigma_a^2Z' + WI\sigma_{c_1}^2W' + HI\sigma_{\partial}^2H' + QI\sigma_r^2Q' + UI\sigma_{c_2}^2U' + I\sigma_e^2.$$

### Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'H & X'Q & X'U \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'H & Z'Q & Z'U \\ W'X & W'Z & W'W + I\lambda_2 & W'H & W'Q & W'U \\ H'X & H'Z & H'W & H'H + I\lambda_3 & H'Q & H'U \\ Q'X & Q'Z & Q'W & Q'H & Q'Q + I\lambda_4 & Q'U \\ U'X & U'Z & U'W & U'H & U'Q & U'U + I\lambda_5 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{\partial} \\ \hat{r} \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ H'y \\ Q'y \\ U'y \end{bmatrix},$$

em que:

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - \partial^2}{h^2};$$

$$\lambda_2 = \frac{\sigma_e^2}{\sigma_{c_1}^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - \partial^2}{c_1^2};$$

$$\lambda_3 = \frac{\sigma_e^2}{\sigma_{\partial}^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - \partial^2}{\partial^2};$$

$$\lambda_4 = \frac{\sigma_e^2}{\sigma_r^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - \partial^2}{r^2};$$

$$\lambda_5 = \frac{\sigma_e^2}{\sigma_{c_2}^2} = \frac{1 - h^2 - c_1^2 - r^2 - c_2^2 - \partial^2}{c_2^2}.$$

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_\partial^2 + \sigma_e^2};$$

$$c_1^2 = \frac{\sigma_{c_1}^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_\partial^2 + \sigma_e^2};$$

$$r^2 = \frac{\sigma_r^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_\partial^2 + \sigma_e^2};$$

$$c_2^2 = \frac{\sigma_{c_2}^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_\partial^2 + \sigma_e^2};$$

$$\partial^2 = \frac{\sigma_\partial^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_\partial^2 + \sigma_e^2}$$

**Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}' Z'y - \hat{c}_1' W'y - \hat{\partial}' H'y - \hat{r}' Q'y - \hat{c}_2' U'y]/[N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \text{tr } (A^{-1} C^{22})]/q$$

$$\hat{\sigma}_{c_1}^2 = [\hat{c}_1' \hat{c}_1 + \hat{\sigma}_e^2 \text{tr } C^{33}]/s_1;$$

$$\hat{\sigma}_\partial^2 = [\hat{\partial}' \hat{\partial} + \hat{\sigma}_e^2 \text{tr } C^{44}]/\eta$$

$$\hat{\sigma}_r^2 = [\hat{r}' \hat{r} + \hat{\sigma}_e^2 \text{tr } C^{55}]/t;$$

$$\hat{\sigma}_{c_2}^2 = [\hat{c}_2' \hat{c}_2 + \hat{\sigma}_e^2 \text{tr } C^{66}]/s_2.$$

**Seqüência de colunas no arquivo de dados para análise no software DFREML**

Indivíduo Pai Mãe Repetição Parcела<sub>1</sub> Bloco Procedência Parcела<sub>2</sub> Variável 1 . . . Variável n

É necessário apenas um arquivo, o qual funciona ao mesmo tempo como

arquivo de pedigree e de dados. Deve ser executado o subprograma DFPREP e em seguida o DFUNI.

O mérito genético total de cada indivíduo é dado pelo somatório do efeito genético aditivo predito com o efeito da procedência a que pertence tal indivíduo.

#### **4.5.2 Avaliação de progêneres de meios irmãos (polinização aberta) de várias populações (procedências), no delineamento em látice, com várias plantas por parcela, várias medições por indivíduo e um só caráter**

Considerando-se as situações (i) (a), (i) (b), (ii) (a) e (ii) (b) descritas no tópico 4.3.1, tem-se que:

- na situação (ii) (a), basta ajustar no vetor de efeitos fixos, os efeitos de procedência e da combinação repetição-medição e adotar o modelo, estimadores e preditores apresentados no tópico 4.4.2.
- na situação (i) (a), basta ajustar no vetor de efeitos fixos, o efeito da combinação repetição-procedência-medição (com bpm níveis, em que p, b e m são os números de procedências, repetições e medições, respectivamente). Neste caso, as observações individuais na medição são ajustadas para a média da procedência na repetição, em uma dada medição. O modelo, estimadores e preditores apresentados no tópico 4.4.2. devem ser empregados.

Para as situações (i)(b) e (ii) (b), excluindo (em caso de não significância) o efeito  $c_2$  em ambas, tem-se os modelos, estimadores e preditores descritos a seguir.

#### **Modelo linear misto (modelo aditivo univariado, multi-populações, de repetibilidade, em látice)**

$y = Xb + Za + Wc_1 + Tp + Qr + H\pi + e$ , em que:

b : vetor de efeitos fixos (no caso, combinações repetição-medição, significando que os valores individuais na medição serão ajustados para a média da repetição na medição);

p : vetor de efeitos permanentes (ambiente permanente dentro de parcela + efeitos genéticos não aditivos);

T : matriz de incidência para p.

## Distribuições e estruturas de médias e variâncias

$$\begin{aligned}
 y & \mid b, V \sim N(Xb, V) \\
 a & \mid A, \sigma_a^2 \sim N(0, A \sigma_a^2) \\
 c_1 & \mid \sigma_{c_1}^2 \sim N(0, I \sigma_{c_1}^2) \\
 \partial & \mid \sigma_\partial^2 \sim N(0, I \sigma_\partial^2) \\
 r & \mid \sigma_r^2 \sim N(0, I \sigma_r^2) \\
 p & \mid \sigma_p^2 \sim N(0, I \sigma_p^2) \\
 e & \mid \sigma_e^2 \sim N(0, I \sigma_e^2)
 \end{aligned}$$

As covariâncias entre todos os efeitos aleatórios do modelo são assumidas como nulas.

Assim:

$$E \begin{bmatrix} y \\ a \\ c_1 \\ p \\ r \\ \partial \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad e \quad Var \begin{bmatrix} a \\ c_1 \\ p \\ r \\ \partial \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_p^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_r^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_\partial^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

$$Var(y) = ZA\sigma_a^2Z' + WI\sigma_{c_1}^2W' + TI\sigma_p^2T' + QI\sigma_r^2Q' + HI\sigma_\partial^2H' + I\sigma_e^2.$$

### Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q & X'H \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'T & Z'Q & Z'H \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'Q & W'H \\ T'X & T'Z & T'W & T'T + I\lambda_3 & T'Q & T'H \\ Q'X & Q'Z & Q'W & Q'T & Q'Q + I\lambda_4 & Q'H \\ H'X & H'Z & H'W & H'T & H'Q & H'H + I\lambda_5 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{p} \\ \hat{r} \\ \hat{\partial} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ Q'y \\ H'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h^2 - c_1^2 - r^2 - \partial^2 - p^2}{h^2};$$

$$\lambda_2 = \frac{\sigma_e^2}{\sigma_{c_1}^2} = \frac{1 - h^2 - c_1^2 - r^2 - \partial^2 - p^2}{c_1^2};$$

$$\lambda_3 = \frac{\sigma_e^2}{\sigma_p^2} = \frac{1 - h^2 - c_1^2 - r^2 - \partial^2 - p^2}{p^2};$$

$$\lambda_4 = \frac{\sigma_e^2}{\sigma_r^2} = \frac{1 - h^2 - c_1^2 - r^2 - \partial^2 - p^2}{r^2};$$

$$\lambda_5 = \frac{\sigma_e^2}{\sigma_\partial^2} = \frac{1 - h^2 - c_1^2 - r^2 - \partial^2 - p^2}{\partial^2}.$$

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_\partial^2 + \sigma_p^2 + \sigma_e^2};$$

$$c_1^2 = \frac{\sigma_{c_1}^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_\partial^2 + \sigma_p^2 + \sigma_e^2};$$

$$r^2 = \frac{\sigma_r^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_\partial^2 + \sigma_p^2 + \sigma_e^2};$$

$$\partial^2 = \frac{\sigma_\partial^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_\partial^2 + \sigma_p^2 + \sigma_e^2};$$

$$p^2 = \frac{\sigma_p^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_\partial^2 + \sigma_p^2 + \sigma_e^2}.$$

Neste modelo, a repetibilidade é dada por

$$\rho = \frac{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_\partial^2 + \sigma_p^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_\partial^2 + \sigma_p^2 + \sigma_e^2}.$$

## **Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}' Z'y - \hat{c}'_1 W'y - \hat{p}' T'y - \hat{r}' Q'y - \hat{\partial}' H'y]/[N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \text{tr } (A^{-1} C^{22})]/q$$

$$\hat{\sigma}_{c_1}^2 = [\hat{c}'_1 \hat{c}_1 + \hat{\sigma}_e^2 \text{tr } C^{33}]/s_1;$$

$$\hat{\sigma}_p^2 = [\hat{p}' \hat{p} + \hat{\sigma}_e^2 \text{tr } C^{44}]/q$$

$$\hat{\sigma}_r^2 = [\hat{r}' \hat{r} + \hat{\sigma}_e^2 \text{tr } C^{55}]/t;$$

$$\hat{\sigma}_{\partial}^2 = [\hat{\partial}' \hat{\partial} + \hat{\sigma}_e^2 \text{tr } C^{66}]/\eta.$$

## **Seqüência de colunas no arquivo de dados para análise no software DFREML**

Indivíduo Pai Mãe Repetição Parcela<sub>1</sub> Permanente Procedência Bloco Variável 1... Variável n

É necessário apenas um arquivo, o qual funciona ao mesmo tempo como de pedigree e de dados. Deve ser executado o subprogramma DFPREP e em seguida o DFUNI.

O mérito genético total de cada indivíduo é dado pelo somatório do efeito genético aditivo predito com o efeito da procedência a que pertence tal indivíduo.

### **4.6 Avaliação apenas de populações**

#### **4.6.1 Avaliação de várias populações (procedências), no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo e um só caráter**

**Modelo linear misto (modelo aditivo univariado, multi-populações, sem parentesco)**

$y = Xb + Za + Wc + Qr + e$ , em que:

$y, b, a, c, r$  e  $e$ : vetores de dados, dos efeitos de blocos (fixos), de efeitos genéticos aditivos dos indivíduos (aleatórios), de efeitos de parcela (aleatórios), de efeitos de procedências (aleatórios) e de erros aleatórios, respectivamente.

$X, Z, W$  e  $Q$ : matrizes de incidência para  $b, a, c, r$  e  $e$ , respectivamente.

### Distribuições e estruturas de médias e variâncias

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

$$a|A, \sigma_a^2 \sim N(0, \sigma_a^2)$$

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

$$r|\sigma_r^2 \sim N(0, I \sigma_r^2)$$

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

$$\text{Cov}(a, c') = 0; \quad \text{Cov}(a, r') = 0; \quad \text{Cov}(a, e') = 0;$$

$$\text{Cov}(r, c') = 0; \quad \text{Cov}(r, e') = 0; \quad \text{Cov}(c, e') = 0$$

ou seja:

$$E \begin{bmatrix} y \\ a \\ c \\ r \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \text{Var} \begin{bmatrix} y \\ a \\ c \\ r \\ e \end{bmatrix} = \begin{bmatrix} V & ZG & WC & Q\mathfrak{R} & R \\ GZ' & G & 0 & 0 & 0 \\ CW' & 0 & C & 0 & 0 \\ \mathfrak{R}Q' & 0 & 0 & \mathfrak{R} & 0 \\ R & 0 & 0 & 0 & R \end{bmatrix}, \text{em que:}$$

$$G = A \sigma_A^2$$

$$R = I \sigma_r^2$$

$$C = I \sigma_c^2$$

$$\mathfrak{R} = I \sigma_e^2$$

$$V = ZA \sigma_a^2 Z' + WI \sigma_c^2 W' + QI \sigma_p^2 Q' + I \sigma_e^2.$$

## Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'Q \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'Q \\ W'X & W'Z & W'W + I\lambda_2 & W'Q \\ Q'X & Q'Z & Q'W & Q'Q + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ Q'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{1-h^2-c^2-r^2}{h^2} = \frac{\sigma_e^2}{\sigma_a^2}; \quad \lambda_2 = \frac{1-h^2-c^2-r^2}{c^2} = \frac{\sigma_e^2}{\sigma_c^2}; \quad \lambda_3 = \frac{1-h^2-c^2-r^2}{r^2} = \frac{\sigma_e^2}{\sigma_r^2}.$$

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_r^2 + \sigma_e^2}$ : herdabilidade individual no sentido restrito no bloco, em uma dada medição;

$r^2 = \frac{\sigma_r^2}{\sigma_a^2 + \sigma_c^2 + \sigma_r^2 + \sigma_e^2}$ : coeficiente de determinação dos efeitos de procedência;

$c^2 = \frac{\sigma_c^2}{\sigma_a^2 + \sigma_c^2 + \sigma_r^2 + \sigma_e^2}$ : correlação devida ao ambiente comum de parcela.

## Estimadores dos componentes de variância por REML via algoritmo EM

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}' Z'y - \hat{c}' W'y - \hat{r}' Q'y] / [N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}' A^{-1} \hat{a} + \hat{\sigma}_e^2 \operatorname{tr}(A^{-1} C^{22})] / q$$

$$\hat{\sigma}_c^2 = [\hat{c}' c + \hat{\sigma}_e^2 \operatorname{tr} C^{33}] / s$$

$$\hat{\sigma}_r^2 = [\hat{r}' \hat{r} + \hat{\sigma}_e^2 \operatorname{tr} C^{44}] / t, \text{ em que:}$$

$C^{22}$ ,  $C^{33}$  e  $C^{44}$  advém de:

$$C^{-1} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{bmatrix}^{-1} = \begin{bmatrix} C^{11} & C^{12} & C^{13} & C^{14} \\ C^{21} & C^{22} & C^{23} & C^{24} \\ C^{31} & C^{32} & C^{33} & C^{34} \\ C^{41} & C^{42} & C^{43} & C^{44} \end{bmatrix}$$

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

**Indivíduo    Pai    Mãe    Bloco    Parcela    Procedência    Variável 1    . . .    Variável n**

É necessário apenas um arquivo, o qual funciona simultaneamente como arquivo de dados e como arquivo de pedigree. Devem ser executados seqüencialmente os subprogramas DFPREP e DFUNI. As colunas referentes a Pai e Mãe devem ser preenchidas com zero, a herdabilidade deve ser conhecida a priori e seu valor fixado no DFREML.

O mérito genético total dos indivíduos é dado pela soma  $a^* = \hat{a} + \hat{r}$ .

## **4.7. Avaliação simultânea de caracteres**

### **4.7.1 Avaliação de progênies de meios irmãos, no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo e vários caracteres, em um só local**

Considerando o caso bivariado, tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

## Distribuições e estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & I\sigma_{c_{12}} & 0 & 0 \\ 0 & 0 & I\sigma_{c_{12}} & I\sigma_{c_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & I\sigma_{e_{12}} \\ 0 & 0 & 0 & 0 & I\sigma_{e_{12}} & I\sigma_{e_2}^2 \end{bmatrix}$$

$$Var(y_1) = V_1 = Z_1 A \sigma_{a_1}^2 Z_1' + W_1 I \sigma_{c_1}^2 W_1' + I \sigma_{e_1}^2$$

$$Var(y_2) = V_2 = Z_2 A \sigma_{a_2}^2 Z_2' + W_2 I \sigma_{c_2}^2 W_2' + I \sigma_{e_2}^2$$

em que:

$\sigma_{a_1}^2$  e  $\sigma_{a_2}^2$  : variâncias genéticas aditivas dos caracteres 1 e 2, respectivamente;

$\sigma_{a_{12}}$  : covariância genética aditiva envolvendo os caracteres 1 e 2, respectivamente;

$\sigma_{c_1}^2$ ,  $\sigma_{c_2}^2$  e  $\sigma_{c_{12}}$  : variâncias entre parcelas, para os caracteres 1 e 2 e covariância entre os 2 caracteres ao nível do efeito de parcela, respectivamente;

$\sigma_{e_1}^2$ ,  $\sigma_{e_2}^2$  e  $\sigma_{e_{12}}$  : variâncias residuais, para os caracteres 1 e 2 e covariância entre os 2 caracteres ao nível do efeito residual, respectivamente.

## Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \end{bmatrix}, \text{ em que:}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}; \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix};$$

$$R^{-1} = R_O^{-1} \otimes I; \quad G^{-1} = G_O^{-1} \otimes A^{-1}; \quad C^{-1} = C_O^{-1} \otimes I;$$

$$R_O = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma_{e_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{11}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c_1}^2 & \sigma_{c_{12}} \\ \sigma_{c_{12}} & \sigma_{c_2}^2 \end{bmatrix};$$

Os componentes de variância estão associados aos parâmetros  $h^2$ ,  $c^2$  e  $r_a$ , da seguinte maneira:

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{c_i}^2 = c_i^2 \sigma_{y_i}^2; \quad \sigma_{e_i}^2 = (1 - h_i^2 - c_i^2) \sigma_{y_i}^2; \quad \sigma_{a_{ij}} = \rho_{a_{ij}} \sigma_{a_i} \sigma_{a_j}, \quad \text{em}$$

que:

$$\rho_{a_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i} \sigma_{a_j}} = \text{correlação genética aditiva entre os caracteres i e j};$$

## Estimadores dos componentes de variância por REML

Nesta situação são recomendados os algoritmos AI (preferencialmente este) e DF. A função geral a ser maximizada é aquela descrita no item 2.

## Seqüência de colunas de dados para análise no software DFREML

**Experimento Indivíduo Pai Mãe Bloco Parcela Variável 1 . . . Variável n**

São necessários dois arquivos, o de dados e um de pedigree, formado pelas colunas **Indivíduo** **Pai** **Mãe**. Devem ser executados os subprogramas DFPREP e DXMUX.

A estimção e predição simultânea de caracteres utiliza os mesmos modelos apresentados nos itens 4.1.3, 4.1.4, 4.2.3, 4.2.4, 4.3.3, 4.3.4, 4.4.3

e 4.4.4, diferindo apenas nas estruturas de variâncias, as quais devem assumir as covariâncias entre caracteres referentes aos efeitos de parcela e residuais, como diferentes de zero.

Outras situações podem ocorrer, como a avaliação simultânea de vários caracteres, em vários locais. Neste caso, as estruturas de variâncias devem assumir as covariâncias referentes aos efeitos de parcela como zero, nas combinações envolvendo caracteres em diferentes locais e, como diferentes de zero nas combinações envolvendo os vários caracteres em um só local. O mesmo comentário aplica-se ao efeito residual.

## 5. Predição de Valores Genéticos e Genotípicos Interpopulacionais (Interespecíficos)

### 5.1. Híbridos envolvendo duas espécies ou populações

Neste caso, o valor genotípico de um híbrido é dado por:

$$G_{(12)} = \mu_{(12)} + a_{(12)} + d_{(12)}$$

$$= \frac{1}{2}(\mu_1 + \mu_2) + h_{(12)} + \frac{1}{2}(a_{12} + a_{21}) + \delta_{(12)} + m_{(12)}, \text{ em que:}$$

$\mu_{(12)}$ : média geral da população híbrida;

$a_{(12)}$ : efeito genético aditivo de um indivíduo híbrido;

$d_{(12)}$ : efeito de dominância de um indivíduo híbrido;

$\mu_1$  e  $\mu_2$ : média geral das populações 1 e 2, respectivamente;

$h_{(12)}$ : heterose no cruzamento entre as populações 1 e 2;

$a_{12}$  e  $a_{21}$ : efeitos genéticos aditivos dos genitores das populações 1 e 2, respectivamente, em cruzamento com a população recíproca;

$\delta_{(12)}$ : efeito de dominância associado à família de irmãos germanos (capacidade específica de combinação entre os genitores das populações 1 e 2);

$m_{(12)}$ : efeito da segregação mendeliana, contemplando uma fração aditiva e outra dominante.

Verificam-se assim, as igualdades:

$$\mu_{(12)} = \frac{1}{2}(\mu_1 + \mu_2) + h_{(12)}$$

$$a_{(12)} + d_{(12)} = \frac{1}{2}(a_{12} + a_{21}) + \delta_{(12)} + m_{(12)}$$

A partir deste modelo geral, quatro situações práticas podem ser consideradas:

- (i) **Predição do comportamento da progênie híbrida com base em informações prévias de  $m_1$ ,  $m_2$ ,  $a_{12}$ ,  $a_{21}$  e  $h_{(12)}$**

Neste caso,  $\hat{G}_{(12)} = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) + \hat{h}_{(12)} + \frac{1}{2}(\hat{a}_{12} + \hat{a}_{21})$ , ou seja, tendo-se

informações sobre as médias das populações, da heterose entre as mesmas e dos efeitos genéticos aditivos interpopulacionais dos genitores a serem cruzados, pode-se predizer a descendência híbrida. Se a correlação entre os efeitos genéticos aditivos interpopulacionais ( $a_{12}$  e  $a_{21}$ ) e intrapopulacionais ( $a_{11}$  e  $a_{22}$ ) for alta, os últimos podem ser utilizados em lugar dos primeiros.

- (ii) **Seleção e recombinação de genitores em programas de seleção recorrente recíproca ou seleção recorrente intrapopulacional nas duas populações**

Neste caso, a seleção dos genitores a serem recombinados deve basear-se no valor genético aditivo interpopulacional dado por:

$$\hat{a}_{(12)}^* = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) + \hat{h}_{(12)} + \hat{a}_{12} = \hat{\mu}_{(12)} + \hat{a}_{12} \text{ na população 1 e}$$

$$\hat{a}_{(21)}^* = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) + \hat{h}_{(12)} + \hat{a}_{21} = \hat{\mu}_{(12)} + \hat{a}_{21} \text{ na população 2.}$$

No caso de alta correlação genética entre os efeitos aditivos interpopulacionais e intrapopulacionais, os últimos ( $\hat{a}_{11}$  e  $\hat{a}_{22}$ ) podem ser usados em lugar da ( $\hat{a}_{12}$  e  $\hat{a}_{21}$ ) e programas de seleção recorrente intrapopulacional podem ser adotados, com eficiência.

- (iii) **Capitalização da capacidade específica de combinação via plantios comerciais por sementes**

Neste caso, torna-se necessária a avaliação de progêniest de irmãos germanos interpopulacionais (híbridos) obtidos sob cruzamentos dialélicos,

fatoriais ou hierárquicos e o valor genotípico predito da descendência no plantio comercial é dado por

$$\begin{aligned}\hat{G}_{(12)} &= \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) + \hat{h}_{(12)} + \frac{1}{2}(\hat{a}_{12} + \hat{a}_{21}) + \hat{\delta}_{(12)} \\ &= \hat{\mu}_{(12)} + \frac{1}{2}(\hat{a}_{12} + \hat{a}_{21}) + \hat{\delta}_{(12)}\end{aligned}$$

#### (iv) Seleção de híbridos superiores para a clonagem

Neste caso, torna-se necessária a avaliação de progêneres de irmãos germanos interpopulacionais, obtidos sob cruzamentos fatoriais, dialélicos ou hierárquicos ou a realização de testes clonais com indivíduos híbridos. Nesta situação, os valores genotípicos totais dos indivíduos são dados por

$$\hat{G}_{(12)} = \hat{\mu}_{(12)} + \hat{a}_{(12)} + \hat{d}_{(12)} .$$

Em geral, dois esquemas de avaliação de progêneres são utilizados:

##### (a) Apenas avaliação de progêneres interpopulacionais

Nesta situação, dois modelos de análise podem ser adotados:

###### **(a.1.) Avaliação integral de $a_{(12)}$ e $d_{(12)}$ a partir de cruzamentos dialélicos, fatoriais ou hierárquicos**

Neste caso, tem-se o seguinte modelo linear misto (**modelo univariado aditivo-dominante**).

$y = Xb + Za_{(12)} + Zd_{(12)} + Wc + e$ , em que:

$y$ ,  $b$ ,  $a_{(12)}$ ,  $d_{(12)}$ ,  $c$ ,  $e$ : vetores de dados, de efeitos de blocos (fixos), de efeitos genéticos aditivos (aleatórios), de efeitos de dominância (aleatórios), de efeitos de parcela (aleatórios) e de erros aleatórios, respectivamente.

$X$ ,  $Z$  e  $W$ : matrizes de incidência para  $b$ ,  $a(d)$  e  $c$ , respectivamente.

#### Estrutura de médias de variâncias

$$E \begin{bmatrix} y \\ a_{(12)} \\ d_{(12)} \\ c \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad Var \begin{bmatrix} a_{(12)} \\ d_{(12)} \\ c \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_{a_{(12)}}^2 & 0 & 0 & 0 \\ 0 & D\sigma_{d_{(12)}}^2 & 0 & 0 \\ 0 & 0 & I\sigma_c^2 & 0 \\ 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

$Var(y) = ZAZ'\sigma_{a_{(12)}}^2 + ZDZ'\sigma_{d_{(12)}}^2 + WIW'\sigma_e^2 + I\sigma_e^2$ , em que:

$\sigma_{a_{(12)}}^2$ : variância genética aditiva interpopulacional;

$\sigma_{d_{(12)}}^2$ : variância de dominância interpopulacional;

$\sigma_c^2$ : variância entre parcelas;

$\sigma_e^2$ : variância ambiental dentro de parcelas;

A e D: matrizes de correlação genética aditiva e de dominância entre os indivíduos em avaliação, respectivamente.

### Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'Z & Z'W \\ Z'X & Z'Z & Z'Z + D^{-1}\lambda_2 & Z'W \\ W'X & W'Z & W'Z & W'W + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a}_{(12)} \\ \hat{d}_{(12)} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_{a_{(12)}}^2} = \frac{1 - h_{a_{(12)}}^2 - c^2}{h_{(12)}^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_{d_{(12)}}^2} = \frac{1 - h_{a_{(12)}}^2 - c^2}{h_{a_{(12)}}^2 - h_{(12)}^2}; \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1 - h_{a_{(12)}}^2 - c^2}{c^2}.$$

$h_{(12)}^2$  e  $h_{a_{(12)}}^2$ : herdabilidades individuais interpopulacionais no sentido restrito e amplo, respectivamente.

$c^2$ : correlação intraclass devida ao ambiente comum da parcela.

### Estimadores iterativos de máxima verossimilhança restrita (REML) pelo algoritmo EM

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}'_{(12)} Z'y - \hat{d}'_{(12)} Z'y - \hat{c}' W'y]/[N - r(x)]$$

$$\hat{\sigma}_{a_{(12)}}^2 = [\hat{a}'_{(12)} A^{-1} \hat{a}_{(12)} + \hat{\sigma}_e^2 \text{tr}(A^{-1} C^{22})]/q$$

$$\hat{\sigma}_{d_{(12)}}^2 = [\hat{d}'_{(12)} D^{-1} \hat{d}_{(12)} + \hat{\sigma}_e^2 \text{tr}(D^{-1} C^{33})]/q$$

$$\hat{\sigma}_c^2 = [\hat{c}'c + \hat{\sigma}_e^2 \text{tr} C^{44}]/s$$

**Seqüência de colunas no arquivo de dados para análise pelo software  
DFREML**

Indivíduo	Pai	Mãe	Bloco	Parcela	Variável 1	...	Variável n
-----------	-----	-----	-------	---------	------------	-----	------------

Este arquivo funciona ao mesmo tempo como arquivo de pedigree e de dados. Devem ser executados seqüencialmente os subprogramas DFPREP e DFUNI.

Exemplo:

Considere a avaliação dos seguintes indivíduos híbridos de *Eucalyptus urophylla* x *E. grandis*, para o caráter altura de plantas aos quatro anos de idade.

Indivíduo	Bloco	Pai ( <i>E. grandis</i> )	Mãe ( <i>E. urophylla</i> )	Altura (metros)
7	1	1	2	25,3
8	2	1	2	22,7
9	1	3	4	14,6
10	2	3	4	17,7
11	1	5	6	19,7

Como existe uma só planta por parcela o modelo passa a ser:

$$y = Xb + Za + Zd + e$$

As matrizes de incidência são:

$$X' = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$Z' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

As matrizes de parentesco A e D são da forma:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & r_1 & r_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & r_1 & r_1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & r_1 & r_1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & r_1 & r_1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & r_1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & r_1 \\ r_1 & r_1 & 0 & 0 & 0 & 0 & 1 & r_2 & 0 & 0 & 0 \\ r_1 & r_1 & 0 & 0 & 0 & 0 & r_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & r_1 & r_1 & 0 & 0 & 0 & 0 & 0 & 1 & r_2 \\ 0 & 0 & r_1 & r_1 & 0 & 0 & 0 & 0 & r_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & r_1 & r_1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

em que  $r_2 = \frac{1}{2}$  em A e  $r_2 = \frac{1}{4}$  em D e  $r_1 = \frac{1}{2}$  em A e  $r_1 = 0$  em D.

As equações de modelo misto sem os efeitos de parcela eqüivalem a:

$$\begin{bmatrix} X'X & X'Z & X'Z \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'Z \\ Z'X & Z'Z & Z'Z + D^{-1}\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a}_{12} \\ \hat{d}_{12} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \end{bmatrix}$$

Assumindo  $\sigma_{a_{(12)}}^2 = 2,75$ ,  $\sigma_{d_{(12)}}^2 = 1,00$  e  $\sigma_e^2 = 3,125$ , tem-se

$h_{(12)}^2 = 0,4$ ,  $h_{a_{(12)}}^2 = 0,55$ ,  $\lambda_1 = 1,1364$  e  $\lambda_2 = 3,125$ . Resolvendo-se as equações de modelo misto, tem-se:

Efeitos	Soluções
<b>Blocos</b>	
$\hat{b}_1$	19,8667
$\hat{b}_2$	20,1803
<b>Efeitos genéticos aditivos</b>	
$\hat{a}_1$	1,2865
$\hat{a}_2$	1,2865
$\hat{a}_3$	-1,2532
$\hat{a}_4$	-1,2532
$\hat{a}_5$	-0,0333
$\hat{a}_6$	-0,0333
$\hat{a}_7$	2,3113
$\hat{a}_8$	1,5482
$\hat{a}_9$	-2,2446
$\hat{a}_{10}$	-1,5149
$\hat{a}_{11}$	-0,0667
<b>Efeitos de dominância</b>	
$\hat{d}_1$	0
$\hat{d}_2$	0
$\hat{d}_3$	0
$\hat{d}_4$	0
$\hat{d}_5$	0
$\hat{d}_6$	0
$\hat{d}_7$	0,7929
$\hat{d}_8$	0,3767
$\hat{d}_9$	-0,7687
$\hat{d}_{10}$	-0,3706
$\hat{d}_{11}$	-0,0242

Verifica-se que somente os efeitos de dominância associados aos indivíduos com observações puderam ser preditos. Para os genitores, somente os efeitos aditivos foram preditos.

De posse das previsões, as quatro situações práticas podem ser consideradas:

### **(i) Predição do comportamento da progênie híbrida**

A média do cruzamento entre os genitores 1 e 6 pode ser predita por:

$$\begin{aligned}
 \hat{G}_{16} &= \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) + \hat{h}_{(12)} + \frac{1}{2}(\hat{a}_1 + \hat{a}_6) \\
 &= \hat{\mu}_{(12)} + \frac{1}{2}(\hat{a}_1 + \hat{a}_6) \\
 &= \frac{19,8667 + 20,1803}{2} + \frac{1}{2}(1,2865 - 0,0333) \\
 &= 20,0235 + 0,6266 \\
 &= 20,6501 \text{ metros}
 \end{aligned}$$

### **(ii) Seleção de genitores para seleção recorrente recíproca**

O valor genético aditivo dos dois melhores genitores de *E. urophylla* equivalem a  $\hat{a}_2^* = \hat{\mu}_{(12)} + \hat{a}_2 = 20,0235 + 1,2865 = 21,3100$  e  $\hat{a}_6^* = \hat{\mu}_{(12)} + \hat{a}_6 = 20,02351 - 0,0333 = 19,9902$ . Assim, recombinando esses indivíduos na população de *E. urophylla*, a contribuição desta população para o ganho no híbrido é dada por  $[(21,3100 + 19,9902)/2]/(20,0235)-1$ , que equivale a 3,13%.

### **(iii) Exploração da capacidade específica de combinação através de plantios por sementes**

O valor genotípico predito da descendência do cruzamento entre os genitores 1 e 2 é dado por  $\hat{G}_{(12)} = \hat{\mu}_{(12)} + \frac{1}{2}(\hat{a}_1 + \hat{a}_2) + \hat{d}_{(12)}$ . No caso,  $\hat{d}_{(12)}$  equivale à média de  $\hat{d}_7$  e  $\hat{d}_8$ , ou seja à média dos efeitos de dominância dos indivíduos que compõem a família 1 x 2. Assim, tem-se:

$$\begin{aligned}
\hat{G}_{(12)} &= \hat{\mu}_{(12)} + \frac{1}{2}(\hat{a}_1 + \hat{a}_2) + \frac{1}{2}(\hat{d}_7 + \hat{d}_8) \\
&= 20,6501 + \frac{1}{2} (1,2865 + 1,2865) + \frac{1}{2} (0,7929 + 0,3767) \\
&= 22,5214 \text{ metros}
\end{aligned}$$

#### (iv) Seleção de indivíduos híbridos superiores para a clonagem

O valor genotípico do melhor indivíduo para clonagem é dado por:

$$\begin{aligned}
\hat{G}_7 &= \hat{\mu}_{(12)} + \hat{a}_7 + \hat{d}_7 \\
&= 20,6501 + 2,3113 + 0,7929 \\
&= 23,7543 \text{ metros}
\end{aligned}$$

#### (a.2.) Avaliação de $a_{(12)}$ e $d_{(12)}$ a partir de cruzamentos dialéticos, fatoriais ou hierárquicos

A construção e inversão da matriz de parentesco de dominância é uma limitação dos programas computacionais disponíveis. Assim, uma forma de contornar este problema é o ajuste do efeito da capacidade específica de combinação ou efeito de dominância comum a uma família de irmãos germanos. Este ajuste não requer o uso da matriz de parentesco de dominância e fornece de maneira mais fácil a predição de  $d_{(12)}$  e a estimativa de  $\sigma_{d_{(12)}}^2$ . Entretanto, não fornece o valor genotípico de cada indivíduo híbrido.

#### Modelo linear misto (modelo univariado de capacidade específica de combinação)

$y = Xb + Za_{(12)} + Wc + Sd_{(12)} + e$ , em que:

$d_{(12)}$  é o efeito aleatório de dominância associado à família de irmãos germanos;

$S$  é a matriz de incidência para  $d_{(12)}$ .

A variância de  $d_{(12)}$  contempla  $(1/4) \sigma_{d_{(12)}}^2$ . Assim,  $\sigma_{d_{(12)}}^2 = 4 \sigma_{\delta_{(12)}}^2$ .

## Estrutura de médias de variâncias

$$E \begin{bmatrix} y \\ a_{(12)} \\ c \\ \delta_{(12)} \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad e \quad Var \begin{bmatrix} a_{(12)} \\ c \\ \delta_{(12)} \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_{a_{(12)}}^2 & 0 & 0 & 0 \\ 0 & I\sigma_c^2 & 0 & 0 \\ 0 & 0 & I(1/4)\sigma_{d_{(12)}}^2 & 0 \\ 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

$Var(y) = ZAZ'\sigma_{a_{(12)}}^2 + WIW'\sigma_c^2 + SIS'(1/4)\sigma_{d_{(12)}}^2 + I\sigma_e^2$ , em que:

## Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'S \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'S \\ W'X & W'Z & W'W + I\lambda_2 & W'S \\ S'X & S'Z & S'W & S'S + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a}_{(12)} \\ \hat{c} \\ \hat{\delta}_{(12)} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ S'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_{a_{(12)}}^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_c^2}; \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_{\delta_{(12)}}^2}.$$

## Estimadores iterativos REML pelo algoritmo EM

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{a}'_{(12)} Z'y - \hat{c}' W'y - \hat{\delta}'_{(12)} S'y]/[N - r(x)]$$

$$\hat{\sigma}_{a_{(12)}}^2 = [\hat{a}'_{(12)} A^{-1} \hat{a}_{(12)} + \hat{\sigma}_e^2 \text{tr}(A^{-1} C^{22})]/q$$

$$\hat{\sigma}_c^2 = [\hat{c}'c + \hat{\sigma}_e^2 \text{tr} C^{33}]/s$$

$\hat{\sigma}_{\delta_{(12)}}^2 = [\hat{\delta}'_{(12)} \hat{\delta}_{(12)} + \hat{\sigma}_e^2 \text{tr} C^{44}]/f$ , em que f é o número de famílias de irmãos germanos.

## Seqüência de colunas no arquivo de dados para análise pelo software DFREML

Indivíduo    Pai    Mãe    Bloco    Parcela    Família    Variável 1 . . .    Variável n

Este arquivo funciona ao mesmo tempo como arquivo de pedigree e de dados. Devem ser executados seqüencialmente os subprogramas DFPREP e DFUNI.

### **(b) Avaliação simultânea de progêniens interpopulacionais e intrapopulacionais**

Muitas vezes são geradas, simultaneamente, progêniens intrapopulacionais e interpopulacionais de cada matriz. Este esquema é vantajoso, pois permite, adicionalmente, a obtenção de informações sobre a correlação genética ( $r_a$ ) entre os efeitos aditivos intrapopulacionais e interpopulacionais e sobre a magnitude da média e variação intrapopulacionais associados a cada população.

Nesta situação, deve-se considerar um modelo multivariado tratando os materiais genéticos intrapopulacionais e interpopulacionais como caracteres distintos. Isto é justificável quando  $r_a$  não equívale a 1.

Quando cruzamentos dialélicos, fatoriais ou hierárquicos são realizados, tem-se a modelagem descrita a seguir.

#### **Modelo linear misto (modelo multivariado de capacidade específica de combinação)**

$$y = Xb + Za + Wc + Sd + e$$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{22} \end{bmatrix} = \begin{bmatrix} X_{11} & 0 & 0 \\ 0 & X_{12} & 0 \\ 0 & 0 & X_{22} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{12} \\ b_{22} \end{bmatrix} + \begin{bmatrix} Z_{11} & 0 & 0 \\ 0 & Z_{12} & 0 \\ 0 & 0 & Z_{22} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{22} \end{bmatrix} + \begin{bmatrix} W_{11} & 0 & 0 \\ 0 & W_{12} & 0 \\ 0 & 0 & W_{22} \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ c_{22} \end{bmatrix} + \begin{bmatrix} S_{11} & 0 & 0 \\ 0 & S_{22} & 0 \\ 0 & 0 & S_{22} \end{bmatrix} \begin{bmatrix} \delta_{11} \\ \delta_{12} \\ \delta_{22} \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{22} \end{bmatrix}$$

$y_{11}, y_{12}, y_{22}$  : vetores de dados correspondentes às progêniens intrapopulacionais da população 1, interpopulacionais do cruzamento 1 x 2 e intrapopulacionais da população 2, respectivamente;

$b_{11}, b_{12}, b_{22}$  : vetores de efeitos fixos (médias dos materiais genéticos nos blocos) associadas aos experimentos avaliando os três conjuntos de materiais genéticos, respectivamente. (De preferência, os três tipos de materiais genéticos devem ser alocados em um mesmo bloco);

$a_{11}, a_{12}, a_{22}$  : vetores aleatórios referentes aos valores genéticos aditivos dos indivíduos na população 1, interpopulação 1 x 2 e população 2, respectivamente;

$c_{11}$ ,  $c_{12}$ ,  $c_{22}$ : vetores aleatórios referentes aos efeitos de parcela associados à avaliação dos três materiais genéticos, respectivamente;

$d_{11}$ ,  $d_{12}$ ,  $d_{22}$ : vetores aleatórios das capacidades específicas de combinação intrapopulacional da população 1, interpopulacional e intrapopulacional da população 2, respectivamente;

$e_{11}$ ,  $e_{12}$ ,  $e_{22}$ : vetores de erros aleatórios associados à avaliação dos três materiais genéticos, respectivamente.

### Estrutura de variâncias

$$Var = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{22} \\ c_{11} \\ c_{12} \\ c_{22} \\ \delta_{11} \\ \delta_{12} \\ \delta_{22} \\ e_{11} \\ e_{12} \\ e_{22} \end{bmatrix} = \begin{bmatrix} A\sigma_{a_{11}}^2 & A\sigma_{a_{112}} & A\sigma_{a_{122}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{112}} & A\sigma_{a_{12}}^2 & A\sigma_{a_{222}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{122}} & A\sigma_{a_{1222}} & A\sigma_{a_{22}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_{11}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{c_{12}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{c_{22}}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{\delta_{11}}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{\delta_{12}}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{\delta_{22}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_{11}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_{12}}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_{22}}^2 \end{bmatrix},$$

em que:

$\sigma_{a_{11}}^2$ ,  $\sigma_{a_{12}}^2$  e  $\sigma_{a_{22}}^2$ : variâncias genéticas aditivas intrapopulacional 1, interpopulacional 1 x 2 e intrapopulacional 2, respectivamente;

$\sigma_{a_{112}}^2$ : covariância genética aditiva entre os efeitos intrapopulacional 1 e interpopulacionais 1 x 2;

$\sigma_{a_{122}}^2 = 0$ : covariância entre os efeitos aditivos intrapopulacionais 1 e 2;

$\sigma_{a_{1222}}^2$ : covariância genética aditiva entre os efeitos interpopulacionais 1 x 2 e os efeitos intrapopulacionais 2;

$\sigma_{\delta_{11}}^2$ ,  $\sigma_{\delta_{12}}^2$  e  $\sigma_{\delta_{22}}^2$ : variâncias dos efeitos da capacidade específica de combinação intrapopulacional 1, interpopulacional 1 x 2 e intrapopulacional 2, respectivamente.

Sendo  $\sigma_{a_{112}} = 0$ , pode-se optar também pelo ajuste de dois modelos bivariados (um envolvendo  $y_{11}$  e  $y_{12}$  e outro envolvendo  $y_{12}$  e  $y_{22}$ ) ao invés do modelo trivariado.

A partir do modelo apresentado, pode-se obter informações sobre a correlação entre os efeitos aditivos intrapopulacionais (capacidade geral de combinação) e efeitos aditivos interpopulacionais (capacidade geral de hibridação) pelas relações  $\rho_{a_{1112}} = \sigma_{a_{1112}} / (\sigma_{a_{11}} \sigma_{a_{12}})$  e  $\rho_{a_{1222}} = \sigma_{a_{1222}} / (\sigma_{a_{12}} \sigma_{a_{22}})$ .

### Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}S \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W & Z'R^{-1}W \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}S \\ S'R^{-1}X & S'R^{-1}Z & S'R^{-1}W & S'R^{-1}S + \Delta^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ S'R^{-1}y \end{bmatrix}, \text{ em}$$

que:

$$R^{-1} = R_o^{-1} \otimes I; \quad G^{-1} = G_o^{-1} \otimes A^{-1}; \quad C^{-1} = C_o^{-1} \otimes I; \quad \Delta^{-1} = \Delta_o^{-1} \otimes I$$

$$R_o = \begin{bmatrix} \sigma_{e_{11}}^2 & 0 & 0 \\ 0 & \sigma_{e_{12}}^2 & 0 \\ 0 & 0 & \sigma_{e_{22}}^2 \end{bmatrix}; \quad G_o = \begin{bmatrix} \sigma_{a_{11}}^2 & \sigma_{a_{1112}}^2 & 0 \\ \sigma_{a_{1112}}^2 & \sigma_{a_{12}}^2 & \sigma_{a_{1222}}^2 \\ 0 & \sigma_{a_{1222}}^2 & \sigma_{a_{22}}^2 \end{bmatrix}; \quad C_o = \begin{bmatrix} \sigma_{c_{11}}^2 & 0 & 0 \\ 0 & \sigma_{c_{12}}^2 & 0 \\ 0 & 0 & \sigma_{c_{22}}^2 \end{bmatrix}; \quad \Delta_o = \begin{bmatrix} \sigma_{\delta_{11}}^2 & 0 & 0 \\ 0 & \sigma_{\delta_{12}}^2 & 0 \\ 0 & 0 & \sigma_{\delta_{22}}^2 \end{bmatrix}$$

### Estimadores dos componentes de variância por REML

Nesta situação, devido à complexidade do modelo e ao elevado número de componentes de variância a serem estimados, são recomendados os algoritmos DF e AI (preferencialmente este). Neste caso, a função geral a ser maximizada é aquela descrita no item 2.

### Seqüência de colunas no arquivo de dados para análise no software DFREML

População Indivíduo Pai Mãe Bloco Parcela Família Variável 1 . . . Variável n

São necessários dois arquivos: um de dados conforme a estrutura apresentada e outro de pedigree contendo apenas as colunas **Indivíduo** **Pai** **Mãe**. Devem ser executados, seqüencialmente, os subprogramas DFPREP e DXMUX.

## 5.2. Híbridos envolvendo três ou mais espécies

Híbridos envolvendo três ou mais espécies têm sido comuns no melhoramento do *Eucalyptus* no Brasil (Assis, 2000). Nesta situação, são envolvidos na análise vários tipos de híbridos. Por exemplo, um híbrido triplo envolvendo *E. urophylla*, *E. dunnii* e *E. grandis* envolve em sua genealogia os híbridos *E. urophylla* x *E. grandis*, *E. dunnii* x *E. grandis* e *E. urophylla* x *E. dunnii*. Também cada indivíduo de uma espécie é envolvido em vários cruzamentos interpopulacionais, de forma que uma avaliação completa deve envolver toda esta mistura de híbridos, procurando identificar os melhores.

Um modelo para esta situação deve contemplar os efeitos fixos da heterose associada a cada tipo de híbrido, por exemplo, efeito de heterose envolvendo as várias combinações de híbridos entre duas espécies e as combinações envolvendo três espécies.

### Modelo linear misto (modelo univariado de heterose fixa)

$y = Xb + Th + Za + Wc + Sd + e$ , em que:

b : vetor de efeitos fixos (médias de blocos);

h : vetor de efeitos fixos da heterose associada a cada tipo de híbrido;

a : vetor de efeitos aditivos interpopulacionais aleatórios, expressos como desvios da média do tipo de híbrido em questão;

c : vetor dos efeitos aleatórios de parcela;

d : vetor aleatório dos efeitos da capacidade específica de combinação entre os dois genitores envolvidos no cruzamento, expressos como desvios da média do tipo de híbrido em questão.

e : vetor de erros aleatórios;

X, T, Z, W e S : matrizes de incidência para b, h, a, c e d, respectivamente.

Neste caso, os valores genotípicos preditos de cada cruzamento visando explorar a capacidade específica de combinação é dado por  $\hat{g} = Xb + Th + Z\bar{a} + S\delta$ , em que  $\bar{a}$  refere-se à média dos efeitos aditivos interpopulacionais dos genitores envolvidos no cruzamento. Indivíduos superiores dentro dos melhores cruzamentos devem ser submetidos a teste clonal.

As equações de modelo misto são similares àquelas descritas no tópico (a.2), bastando incluir o efeito fixo da heterose, fato que requer uma coluna a mais no arquivo de dados, codificando os vários tipos de híbridos.

## 6. Testes Clonais

### 6.1 Avaliação de clones não aparentados no delineamento em blocos ao acaso com várias plantas por parcela (válido também para linhagens não aparentadas de espécies autógamas)

**Modelo linear misto (modelo univariado de clones repetidos)**

$y = Xb + Zg + Wc + e$ , em que

$y, b, g, c \text{ e } e$ : vetores de dados, de efeitos fixos (blocos), de efeitos genotípicos de clones (aleatórios), de efeitos de parcela (aleatórios) e de erros aleatórios, respectivamente.

$X, Z \text{ e } W$ : matrizes de incidência para  $b, g \text{ e } c$ , respectivamente.

**Distribuições e estruturas de médias e variâncias**

$$E \begin{bmatrix} y \\ g \\ c \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} g \\ c \\ e \end{bmatrix} = \begin{bmatrix} I\sigma_g^2 & 0 & 0 \\ 0 & I\sigma_c^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

**Equações de modelo misto**

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + I\lambda_1 & Z'W \\ W'X & W'Z & W'W + I\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_g^2} = \frac{1 - h_a^2 - c^2}{h_a^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1 - h_a^2 - c^2}{c^2}.$$

$$h_a^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_c^2 + \sigma_e^2} : \text{herdabilidade individual no sentido amplo no bloco};$$

$\sigma_g^2$ : variância genotípica entre clones;.

$\sigma_c^2$ : variância entre parcelas;

$\sigma_e^2$ : variância residual ou ambiental dentro de parcelas.

### **Estimadores iterativos de componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y'y - \hat{b}' X'y - \hat{g}' Z'y - \hat{c}' W'y] / [N - r(x)]$$

$$\hat{\sigma}_g^2 = [\hat{g}'\hat{g} + \hat{\sigma}_e^2 \operatorname{tr} C^{22})] / q$$

$$\hat{\sigma}_c^2 = [\hat{c}'c + \hat{\sigma}_e^2 \operatorname{tr} C^{33}] / s$$

### **Estimação e predição pelo software DFREML**

O modelo de estimação e predição apresentado é similar a um modelo de genitor (“sire model”) e não a um modelo individual (“animal model”). Como o software DFREML trabalha com um modelo individual, para análises de testes clonais, tal programa pode ser utilizado ajustando um modelo de repetibilidade. Neste caso, a variância de “ambiente permanente” contemplará a variância não aditiva total livre do componente de ambiente permanente (já que as medidas repetidas em um mesmo clone são realizadas em diferentes ramegos), e o efeito de “ambiente permanente” contemplará os efeitos não aditivos. Assim, para obtenção dos efeitos genotípicos totais (aditivo + não aditivo) de cada clone basta somar os efeitos aditivos e de “ambiente permanente” ajustados pelo programa. Este procedimento é válido quando os clones não são aparentados. A forma do arquivo para este tipo de análise é apresentada a seguir:

Indivíduo	Pai	Mãe	Bloco	Permanente	Parcela	Vetor de Dados
-----------	-----	-----	-------	------------	---------	----------------

Neste caso, o programa ajustará os efeitos aditivos, permanente de dominância (aleatório adicional), de parcela (aleatório adicional) e fixo de bloco.

No caso de testes clonais instalados no delineamento em látice, basta substituir o efeito fixo de bloco pelo efeito de repetições e incluir no modelo, o efeito aleatório de blocos dentro de repetições.

## 6.2 Avaliação de clones aparentados no delineamento em blocos ao acaso, com várias plantas por parcela

**Modelo linear misto (modelo univariado aditivo-dominante)**

$$y = Xb + Za + Zd + Wc + e, \text{ em que}$$

d: vetor aleatório dos efeitos de dominância.

**Distribuições e estruturas de médias e variâncias**

$$E \begin{bmatrix} y \\ a \\ d \\ c \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} y \\ a \\ d \\ c \\ e \end{bmatrix} = \begin{bmatrix} V & ZA\sigma_a^2 & ZD\sigma_d^2 & WI\sigma_c^2 & I\sigma_e^2 \\ A\sigma_a^2 Z' & A\sigma_a^2 & 0 & 0 & 0 \\ D\sigma_d^2 Z' & 0 & D\sigma_d^2 & 0 & 0 \\ I\sigma_c^2 W' & 0 & 0 & I\sigma_c^2 & 0 \\ I\sigma_e^2 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}, \text{ em que:}$$

$$V = ZA\sigma_a^2 Z' + ZD\sigma_d^2 Z' + WI\sigma_c^2 W' + I\sigma_e^2.$$

**Equações de modelo misto**

$$\begin{bmatrix} X'X & X'Z & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'Z & Z'W \\ Z'X & Z'Z & Z'Z + D^{-1}\lambda_2 & Z'W \\ W'X & W'Z & W'Z & W'W + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{d} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h_a^2-c^2}{h^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_d^2} = \frac{1-h_a^2-c^2}{h_a^2-h^2}; \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1-h_a^2-c^2}{c^2}.$$

$\sigma_d^2$  e  $h_a^2$ : variância genética de dominância e herdabilidade no sentido amplo, respectivamente;

D : matriz de correlação genética de dominância entre os indivíduos em avaliação.

O sistema apresentado prediz isoladamente os efeitos aditivos ( $\hat{a}$ ) e de dominância ( $\hat{d}$ ). Os valores genotípicos totais, dados por  $\hat{g} = \hat{a} + \hat{d}$ , podem ser preditos diretamente pelas equações de modelo misto:

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + G^{-1}\sigma_e^2 & Z'W \\ W'X & W'Z & W'W + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$G = A\sigma_a^2 + D\sigma_d^2$$

### **Estimadores iterativos dos componentes de variância por REML via algoritmo EM**

$$\hat{\sigma}_e^2 = [y'y - \hat{b}'X'y - \hat{a}'Z'y - \hat{d}'Z'y - \hat{c}'W'y]/[N - r(x)]$$

$$\hat{\sigma}_c^2 = [\hat{c}'\hat{c} + \hat{\sigma}_e^2 \operatorname{tr} C^{44}]/s;$$

$$\hat{\sigma}_a^2 = [\hat{a}'A^{-1}\hat{a} + \hat{\sigma}_e^2 \operatorname{tr} (A^{-1}C^{22})]/q;$$

$$\hat{\sigma}_d^2 = [\hat{d}'D^{-1}\hat{d} + \hat{\sigma}_e^2 \operatorname{tr} (D^{-1}C^{33})]/q.$$

### **Seqüência de colunas no arquivo de dados para análise no software DFREML**

**Indivíduo      Pai      Mãe      Bloco      Parcela      Variável 1      . . .      Variável n**

Para ajuste dos efeitos de dominância, os mesmos devem ser considerados como “segundo efeito aleatório por indivíduo”. Neste caso, além dos arquivos de dados e de pedigree, deve ser fornecido um arquivo adicional com nome padrão DF45#DAT, referente à inversa da matriz de parentesco de dominância. Este arquivo não formatado deve fornecer todos os elementos não zero do triângulo inferior da matriz inversa e deve conter três colunas: um código inteiro (de 1 ao número de indivíduos na análise) referente ao número da coluna na matriz; um código inteiro referente ao número da linha (maior ou igual ao número da coluna) na matriz (de 1 ao número de indivíduos na análise); uma variável verdadeira (real) fornecendo o elemento da inversa da matriz de parentesco de dominância.

## 7. Modelos não Lineares para Variáveis Binomiais e Categóricas

Em teoria, variáveis categóricas e binomiais (tais como a presença ou ausência de determinados atributos nos indivíduos) não são bem descritas por modelos estatísticos lineares. Para estas variáveis, os modelos não lineares podem ser mais apropriados.

A técnica de modelos lineares generalizados (GLM), desenvolvida por Nelder & Wederburn (1972), permite a generalização ou flexibilização dos modelos lineares clássicos de variáveis contínuas, de forma que toda a estrutura para a estimativa e predição em modelos lineares normais, pode ser estendida para os modelos não lineares. Os modelos lineares clássicos são, em verdade, casos especiais de modelos lineares generalizados.

Neste caso, a variável observacional pode ser definida pelo modelo  $y = \mu + e$ , em que  $\mu$  refere-se à esperança de  $y$  e  $e$  refere-se ao vetor de erros aleatórios.

No caso especial em que  $y$  segue uma distribuição normal e  $g(\mu)$  é uma ligação identidade, obtém-se o tradicional modelo linear misto  $y = Xb + Za + e$ ,

em que  $E(y) = Xb$ ,  $Cov(a) = G$ ,  $Cov(e) = R = I\sigma_e^2$  e, consequentemente,

$$Cov(y) = ZGZ' + R = ZGZ' + I\sigma_e^2.$$

Para o caso em que  $y$  segue uma distribuição binomial, tem-se o caso da estimativa e predição em modelos lineares generalizados com efeitos fixos e aleatórios, conforme Schall (1991).

A função de ligação logito,  $g(\mu)$ , aplicada aos dados  $y$  é linearizada, conforme a expansão em série de Taylor de primeira ordem fornecendo  $y^*$ , da seguinte forma:

$$y^* = g(y) = g(\mu) + (y - \mu) g'(\mu)$$

Assim tem-se:

$$\begin{aligned} y_i^* &= \eta_i + \frac{y_i - \mu_i}{\mu_i(1-\mu_i)} \\ &= \log\left(\frac{\mu_i}{1-\mu_i}\right) + \frac{y_i - \mu_i}{\mu_i(1-\mu_i)}. \end{aligned}$$

De posse da variável observacional (ou dependente) ajustada  $y^*$ , tem-se que o modelo linear misto equivale a  $y^* = Xb + Za + eg'(\mu)$ , em que:

$$E(y^*) = Xb, \ Cov(a) = G, \ Cov[eg'(\mu)] = W^{-1}\sigma_e^2 \ e \ Cov(y^*) = ZGZ' + W^{-1}\sigma_e^2.$$

O modelo  $y^* = Xb + Za + e$  tem a mesma estrutura de primeira e segunda ordem que o modelo  $y = Xb + Za + e$  de forma que os algoritmos de estimação e predição para o caso normal podem ser adaptados, apenas substituindo  $y$  por  $y^*$  e  $Cov(e) = R$  por  $Cov[eg'(\mu)] = W^{-1}\sigma_e^2$ .

Assim, tem-se a seguintes equações de modelo misto:

$$\begin{bmatrix} X'S^{-1}X & X'S^{-1}Z \\ Z'S^{-1}X & Z'S^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b}_L \\ \hat{a}_L \end{bmatrix} = \begin{bmatrix} X'S^{-1}y^* \\ Z'S^{-1}y^* \end{bmatrix}, \text{ em que:}$$

$S^{-1}$  = matriz com termos diagonais dados por  $\mu_i (1 - \mu_i) \frac{1}{\sigma_{e_L}^2}$ ;

$\sigma_{e_L}^2$  = variância residual na escala contínua de tolerância (“liability”);

$b_L$  e  $a_L$  = efeitos fixos e aleatórios na escala de tolerância.

No caso em que a refere-se a um vetor de valores genéticos aditivos, tem-se que  $Cov(a) = G = A\sigma_{a_L}^2$ , em que  $A$  é a matriz de correlação genética aditiva entre os indivíduos e  $\sigma_{a_L}^2$  é a variância de  $a_L$ . Neste caso, os estimadores REML são dados por:

$$\hat{\sigma}_{a_L}^2 = \frac{\hat{a}'_L A^{-1} \hat{a}_L}{q - \text{tr}(A^{-1} C^{22})/\sigma_{a_L}^2}; \quad \hat{\sigma}_{e_L}^2 = \frac{(y - X\hat{b}_L - Z\hat{a}_L)' S^{-1} (y - X\hat{b}_L - Z\hat{a}_L)}{N - r(x) - q + \text{tr}(A^{-1} C^{22})/\sigma_{a_L}^2} \sigma_{e_L}^2$$

Em resumo, o processo de estimação envolve:

- (a) estimação de  $\mu = n_1 / N$ , em que  $n_1$  é o número de indivíduos que recebem o escore 1, dentre  $N$  indivíduos avaliados;
- (b) obtenção de  $y^*$ , a partir de  $y$  e  $m$  (neste passo, a variável passa do intervalo  $(0, 1)$  para a reta real, ou seja, a função é linearizada);
- (c) estimação de  $b_L$  e  $\hat{a}_L$ , dados os valores atuais ou correntes de  $\mu$ ,  $\sigma_{e_L}^2$  e  $\sigma_{a_L}^2$ ;

- (d) obtenção de  $\hat{\sigma}_{e_L}^2$  e  $\hat{\sigma}_{a_L}^2$  iterativamente e, após a convergência, proceder à obtenção atualizada de  $\hat{b}_L$  e  $\hat{a}_L$ ;
- (e) obtenção de  $\hat{\eta} = \hat{\theta} = X\hat{b}_L + Z\hat{a}_L$ ;
- (f) obtenção de novo valor predito de  $m$ , usando a função de ligação, através de  $\hat{\mu}_L = \frac{e^{\hat{\theta}}}{1+e^{\hat{\theta}}}$  (neste passo, a variável volta ao intervalo  $(0, 1)$ );
- (g) atualização de  $S^{-1}$  via  $S^{-1} = \hat{\mu}_1(1-\hat{\mu}_1) \frac{1}{\hat{\sigma}_{eL_1}^2}$  e de  $y^*$  via  $y^* = \log \left( \frac{\hat{\mu}_1}{1-\hat{\mu}_1} \right) + \frac{y - \hat{\mu}_1}{\hat{\mu}_1(1-\hat{\mu}_1)} = \hat{\theta} + \frac{y - \hat{\mu}_1}{\hat{\mu}_1(1-\hat{\mu}_1)}$ ;
- (h) voltar ao passo (c), enquanto não se atingir a convergência.

É interessante notar que este algoritmo é essencialmente hierárquico, havendo, a cada iteração compreendendo os passos de (a) a (h), a necessidade de convergência no passo (d).

### Software

Dentre os softwares mencionados neste artigo, o ASREML é o único adequado para a análise de variáveis binomiais. Para a versão "Windows", o arquivo executável necessário é o ASRWIN.EXE. Para utilização do referido aplicativo na análise de modelos mistos ao nível de plantas individuais, são necessários 3 arquivos: um arquivo de dados, um arquivo de pedigree (com os números de identificação dos genitores precedendo os números de identificação dos descendentes) e um arquivo de comandos (este contendo o modelo de análise, o qual deve ser escrito pelo usuário). Os arquivos devem ser preparados em formato ASCII, sendo que os arquivos de dados e de pedigree podem ser salvos usando o programa Notepad (ou bloco de notas, usando documento texto, ou seja extensão .txt) ou mesmo o EXCEL usando a opção de salvar como "Texto (OS/2 ou MS-DOS)", o que produz uma extensão .txt.

O arquivo de comandos deve conter a extensão .as e deve ser escrito com base em 5 seções: (i) linha de título; (ii) definição das colunas do arquivo

de dados; (iii) definição dos arquivos de pedigree e de dados; (iv) definição do modelo estatístico; (v) definição do modelo de variância (esta seção não é necessária em alguns casos). Este arquivo com extensão .as pode ser composto e salvo facilmente dentro do próprio ASREML, através da modificação de arquivos pré-existentes. Os arquivos de resultados mais importantes do programa são aqueles com extensão .sln (o qual apresenta a solução para todos os efeitos do modelo com seus respectivos desvios padrões) e .asr (o qual sumariza os dados e a seqüência das iterações, apresenta as estimativas dos componentes de variância, a análise de variância para os efeitos fixos e suas soluções).

Para a análise de variáveis binomiais, considere como exemplo um experimento instalado no delineamento em blocos ao acaso, com 33 progênieis de meios irmãos (polinização aberta) e 6 blocos, em que foi avaliada a variável sobrevivência (Sob). Tendo-se salvos os arquivos de dados e de pedigree com os nomes Dados.txt e Pedigree.txt, tem-se o seguinte arquivo com extensão .as :

```
Titulo  
Individuo !P  
Pai  
Mae 33  
Bloco 6  
Sob  
Pedigree.txt !SKIP 1 !MAKE  
Dados.txt !SKIP 1  
Sob !BIN !LOGIT ~ Bloco !r Individuo
```

Este arquivo de comando possui a seguinte interpretação:

- (i) Linha 1 : Título qualquer;
- (ii) Linhas 2 a 6 : Identificação de colunas no arquivo de dados, contendo os respectivos números de níveis;
- (iii) Linhas 7 e 8 : Identificação dos arquivos de pedigree e de dados;
- (iv) Linha 9 : Especificação do modelo de análise para a variável Sob;

Ainda no arquivo com extensão .as, tem-se que os comandos advêm após o símbolo !. Assim, tem-se:

- P:** indica que o número de indivíduos deve ser lido no arquivo de pedigree;
- SKIP:** indica que deve ser ignorada a primeira linha dos arquivos, pois refere-se apenas às identificações;
- MAKE:** indica que deve ser feita a matriz de parentesco;
- BIN:** indica que a variável apresenta distribuição binomial;
- LOGIT:** indica que deverá ser usada a função de ligação logito;
- r Individuo:** indica que os efeitos de indivíduos são aleatórios;
- ~ Bloco:** indica que os efeitos de blocos são fixos.

No caso, a seção (v) não foi necessária. Outra opção de análise refere-se à adoção de um modelo de genitor, sendo preditos (1/2) dos efeitos genéticos dos genitores e estimado (1/4) da variação genética aditiva. Neste caso, não é necessário o arquivo de pedigree e o arquivo .as apresenta como conteúdo:

Titulo  
Individuo 1188  
Mae 33  
Bloco 6  
Sob  
Dados.txt !SKIP 1  
Sob !BIN !LOGIT ~ Bloco !r Mae

O número 1188 refere-se ao número total de indivíduos, considerando que havia 6 plantas por parcela.

A seguir será ilustrada a aplicação do software para a análise de uma variável contínua, como o peso de frutos em cacau. Considerando a situação descrita no item 4.6 (cruzamentos dialélicos com medidas repetidas em cada indivíduo), tem-se o seguinte arquivo .as, considerando cruzamentos envolvendo 5 mães e 5 pais, experimentação em 4 blocos, 4 medições por indivíduo, 10 famílias de irmãos germanos, 80 parcelas e 1200 indivíduos no total:

```
Título
Individuo !P
Pai 5
Mae 5
Bloco 4
Medicao 4
Familia 10
Permanente 1200
Parcela 80
Peso
Pedigree.txt !SKIP 1 !MAKE !REPEAT
Dados.txt !SKIP 1
Peso ~ Bloco Medicao !r Individuo Familia Permanente Parcela
```

No caso, os efeitos de bloco e medição foram ajustados como fixos. O comando REPEAT indica que se trata de um caso de medidas repetidas.

Após compostos os arquivos de dados, de pedigree e de comandos, basta executar este último e, então, abrir os arquivos de resultados.

As variáveis categóricas são, provavelmente, uma aproximação da variável real de interesse, sendo que, muitas vezes, as categorias surgem porque não é possível medir a variável real de interesse. Tomar as variáveis categóricas como normais é tanto mais apropriada quanto mais normais forem os escores. Assim, quanto maior o número de categorias, menor é a relevância da transformação das variáveis ou dos modelos para se adequarem as variáveis.

Considere a avaliação da variável número de frutos (denominada Frutos) de cacaueiros em dois locais, no delineamento em blocos ao acaso com 18 progênies de meios irmãos e cinco blocos em cada local. Assumindo normalidade, tem-se que a composição do arquivo de comandos .as para a análise do modelo bivariado, equívale a:

```

Titulo
Individuo !P
Pai
Mae 18
Bloco 10
Parcela 180
Frutos1 !M 0
Frutos2 !M 0
Pedigree.txt !SKIP 1 !MAKE
Dados.txt !SKIP 1
!ASUV
Frutos1 Frutos2 ~ Trait Tr.Bloco !r Tr.Individuo Tr.Parcela !f mv
2 1 2
0 !s2=ve1
0 !s2=ve2

Tr.Individuo 2
Tr 0 US va1 va12 va2
Individuo

Tr.Parcela 2
Tr 0 US vc1 vc2
Parcela

!end

```

O comando !MO deve ser incluído visando converter os valores zero do arquivo em valores inexistentes ou perdidos. No arquivo, os dados referentes a cada local devem ser colocados em duas colunas distintas, preenchendo-se os dados inexistentes com zero (neste caso).

Por outro lado, o comando !ASUV é usado quando os dados são apresentados em uma forma multivariada mas a análise requerida refere-se a um único caráter. Com esta opção, se existem valores perdidos no arquivo de dados, deve-se incluir o comando !f mv no final da linha do modelo linear. O comando !ASUV deve ser colocado em uma linha logo após a linha de denominação do arquivo de dados e antes da linha referente ao modelo linear.

As linhas após o modelo linear referem-se à definição do modelo (estrutura) de variância. No caso, ve<sub>1</sub> e ve<sub>2</sub> referem-se aos valores iniciais para as variâncias residuais nos locais 1 e 2, respectivamente, e vc<sub>1</sub> e vc<sub>2</sub> referem-se aos valores

iniciais para as variâncias entre parcelas nos locais 1 e 2, respectivamente. Por outro lado,  $va_1$ ,  $va_2$  e  $va_{12}$ , referem-se aos valores iniciais para as variâncias genéticas aditivas nos locais 1 e 2 e covariância genética aditiva entre os locais 1 e 2, respectivamente.

## 8. Ajuste de Covariável e Análise de Covariância

O ajuste de covariáveis é importante como forma de reduzir o erro experimental, através da eliminação de certas diferenças ambientais aleatórias entre parcelas dentro de blocos, como por exemplo, a sobrevivência diferenciada entre parcelas, não devidas a causas genéticas. Nesta situação, o número de plantas na parcela deve ser ajustado como uma covariável. Outro exemplo, em que o ajuste de uma covariável pode ser necessário, refere-se à situação em que os indivíduos são avaliados em diferentes idades, para uma determinada característica. Neste caso, a idade deve ser ajustada ao modelo, como uma covariável.

Considerando um modelo estatístico tradicional e incluindo como uma covariável  $x$ , o número de plantas na parcela, tem-se (Ramalho et al., 2000):

$$Y_{ij} = m + p_i + r_j + b(x_{ij} - \bar{x}) + e_{ij}, \text{ em que:}$$

$Y_{ij}$ : observação da variável dependente de interesse, referente à progênie i no bloco j.;

$m$ ,  $p_i$ ,  $r_j$ ,  $e$ : efeitos da média geral, da progênie i, do bloco j e do erro experimental, respectivamente;

$x_{ij}$ : número de plantas na parcela ij (variável dependente), com média  $\bar{x}$ ;

$b$ : coeficiente de regressão linear entre x e y;

Verifica-se que o componente  $b(x_{ij} - \bar{x})$  estaria inflacionando o erro experimental, caso não fosse realizado o ajuste da covariável.

Dentre os requisitos para uso da análise de covariância ou ajuste de uma covariável, citam-se, conforme Steel e Torrie (1980) e Ramalho et al. (2000):

- (i) As covariáveis x são efeitos fixos, medidos sem erros e independentes dos efeitos de tratamentos (progênies). Isto implica que não pode haver diferenças significativas entre progênies para a covariável x;
- (ii) A regressão de x em y, após a remoção das diferenças entre blocos e tratamentos, é linear e independente dos tratamentos e blocos. Isto significa que o efeito de x é no sentido de aumentar ou diminuir y por uma constante (b) multiplicada por  $(x_{ij} - \bar{x})$ ;
- (iii) O erro experimental possui distribuição normal com média zero e variância  $I\sigma_e^2$ .

Dessa forma, o ajuste do número de plantas por parcela como covariável na metodologia de modelos mistos somente é recomendável quando a sobrevivência no experimento não apresentar controle genético, ou seja, quando a mesma ocorrer de forma aleatória. Para verificar esta suposição, recomenda-se a análise de variância e a verificação da significância da fonte de variação progêniens. A significância desta fonte de variação indica que existem diferenças genéticas entre progêniens para o caráter e, portanto, o ajuste do número de plantas como covariável não é recomendado. Neste caso, é melhor realizar a predição de valores genéticos para a sobrevivência e posteriormente utilizá-la também na seleção, conforme Resende (1999).

No contexto dos modelos mistos, a verificação do controle genético da sobrevivência pode ser realizada utilizando-se os próprios valores estimados da herdabilidade e de seu desvio padrão. Embora a não significância das diferenças de sobrevivência entre progêniens, signifique que tal variável não está afetando a comparação das mesmas para o outro caráter de interesse (variável dependente), o ajuste da covariável pode ser relevante por permitir um melhor ajuste dos dados de plantas individuais e, por conseguinte, aumentar a acurácia da predição dos valores genéticos individuais. Logicamente, o efeito benéfico é tanto maior quanto maior for a variação na covariável e maior for a associação entre x e y.

Na metodologia de modelos mistos, o ajuste de covariáveis é realizado previamente à estimação e predição. Assim, os modelos, estimadores e preditores apresentados independem do ajuste ou não de covariáveis e são válidos em geral. Para análise pelo software DFREML, basta acrescentar, no arquivo de dados, antes da primeira coluna de dados (primeiro caráter), uma coluna referente à covariável. No caso da sobrevivência, esta coluna contemplaria o número de plantas na parcela.

Finalmente, é importante ressaltar que o ajuste da covariável não é totalmente suficiente para considerar o efeito da competição diferenciada devida às falhas. Isto, porque o estande da parcela não considera as posições das falhas.

## 9. Espécies com Sistema Reprodutivo Misto

Os modelos, estimadores e preditores apresentados nos itens 4.1.1, 4.1.2, 4.1.3, 4.1.4, 4.3.1, 4.3.2, 4.3.3, 4.3.4, 4.4.1, 4.4.2, 4.4.3, 4.4.4, 4.5.1 e 4.5.2 são adequados também para populações com sistema reprodutivo misto. Neste caso, as herdabilidades estimadas devem ser corrigidas levando-se em consideração a taxa de autofecundação e os coeficientes de parentesco corrigidos apresentados por Resende et al. (1995). Posteriormente, as herdabilidades (ou variâncias e covariâncias genéticas) corrigidas devem ser

fixadas nos programas computacionais, visando à predição dos valores genéticos.

## 10. Referências Bibliográficas

- ASSIS, T.F. Production and use of *Eucalyptus* hybrids for industrial purposes. In: DUNGEY, H.S.; DIETERS, M.J.; NIKLES, D.G., ed. **Hybrid breeding and genetics of forest trees**: QFRI/CRC-SPF Symposium Noosa, Queensland, Australia, 9-14 April 2000. Brisbane: Department of Primary Industries, 2000. p.63-74.
- BOLDMAN, K.G.; KRIESE, L.A.; VLECK, L.D. van.; KACHMAN, S.D. **A manual for use of MTDFREML**: a set of programs to obtain estimates of variances and covariances. Washington: ARS / USDA, 1995. 120p.
- BUENO FILHO, J.S.S. **Seleção combinada versus seleção sequencial no melhoramento de populações florestais**. Piracicaba: ESALQ/USP, 1992, 96p. Tese Mestrado.
- BUENO FILHO, J.S.S. **Uso de modelos mistos no melhoramento de populações florestais**. Piracicaba: ESALQ/USP, 1997, 160p. Tese Doutorado.
- COSTA, R.B.; RESENDE, M.D.V. de; ARAÚJO, A J.; GONÇALVES, P. S. BLUP individual e índice multi-efeitos aplicados ao melhoramento da seringueira. In: SIMPÓSIO SOBRE RECURSOS GENÉTICOS PARA A AMÉRICA LATINA E CARIBE, 2., 1999, Brasília. Brasília, **Anais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 1999. 1 CD ROM.
- CRUZ, C.D. **Programa Genes**: aplicativo computacional em Genética e Estatística. Viçosa: Universidade Federal de Viçosa, 1997. 422p.
- CRUZ, C.D.; REGAZZI, O.J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: Universidade Federal de Viçosa, 1994. 390p.
- CULLIS, B.R.; GOGEL, B.; VERBYLA, A.P.; THOMPSON, R. Spacial analysis of multi-environment early generation variety trials. **Biometrics**, v.54, p.1-18, 1998.
- DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistic Society**, London, v.39, p.1-38, 1977.
- FARIAS NETO, J.T.; RESENDE, M.D.V. de. Predição de valores genéticos e estimativa de componentes de variância em pupunheira (*Bactris gasipaes*) pela metodologia BLUP/REML com ênfase na produtividade de palmito. **Genetics and Molecular Biology** , v.23, n.3, Suppl., p.533, 2000.

- GILMOUR, A.R.; CULLIS, B.R.; WELHAM, S.J. THOMPSON, R. **ASREML reference manual**. Orange: NSW Agriculture, 2000. 218p.
- GILMOUR, A.R.; THOMPSON, R.; CULLIS, B.R. Average information REML: An efficient algorithm for parameter estimation in linear mixed models. **Biometrics**, v.51, p.1440-1450, 1995.
- GRASER, H.U.; SMITH, S.P.; TIER, B. A derivative free approach for estimating variance components in animal models by restricted maximum likelihood. **Journal of Animal Science**, Champaign, v.64, n.5, p.1362-1370, 1987.
- HARVILLE, D.A.; CARRIQUIRY, A.L. Classical and Bayesian prediction as applied to unbalanced mixed linear models. **Biometrics**, v.48, p.987-1003, 1992.
- HENDERSON, C.R. **Aplications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462p.
- HENDERSON, C.R.; QUAAS, R.L. Multiple trait evaluation using relatives'records. **Journal of Animal Science**, v.43, p.1188-1197, 1976.
- JOHNSON, D.L.; THOMPSON, R. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. **Journal of Dairy Science**, v.78, p.449-456, 1995.
- KALIL, A.N.; RESENDE, M.D.V.; KALIL, G.P.C. Componentes de variância e predição de valores genéticos em seringueira pela metodologia de modelos mistos (REML/BLUP). **Pesquisa Agropecuária Brasileira**, v.35, n.9, p.1883-1887, 2000.
- MEYER, K. DFREML – a set of programs to estimate variance components under an individual animal model. **Journal of Dairy Science**, v.2, Suppl, p.33-34, 1988.
- MEYER, K. **DFREML: version 3.0b user notes**. Edinburgh: Institute of Animal Genetics of Edinburgh / Armidale: University of New England, Animal Genetics and Breeding Unit, 1998. 31p.
- MEYER, K. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. **Genetique, Selection, Evolution**, Paris, v.23, p.67-83, 1991.
- MEYER, K. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. **Genetics, Selection, Evolution**, v.21, p.317-340, 1989.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistic Society, Series A**, v.135, p.370-384, 1972.

- PAIVA, J.R.; RESENDE, M.D.V. de; CORDEIRO, E.R. Índice multi-efeitos (BLUP) e estimativas de parâmetros genéticos aplicados ao melhoramento da acerola. **Pesquisa Agropecuária Brasileira**, 2000. no prelo.
- PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v.58, p.545-554, 1971.
- PIRES, I.E.; CRUZ, C.D.; BORGES, R.C.G.; REGAZZI, A.J. Índice de seleção combinada aplicada ao melhoramento de *Eucalyptus* spp. **Revista Árvore**, v.20, n.2, p.191-198, 1996.
- RAMALHO, M.A.P.; SANTOS, J.B.; ZIMMERMANN, M.J.O. **Genética quantitativa no melhoramento de plantas autógamas**. Goiânia: Universidade Federal de Goiás, 1993.
- RAMALHO, M.A.P.; FERREIRA, D.F.; OLIVEIRA, A.C. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, 2000. 303p.
- RESENDE, M.D.V. de. Genética biométrica no melhoramento de fruteiras. In: ENCONTRO DE GENÉTICA DO NORDESTE, 15., 2000. **Anais on line...** Fortaleza, 2000. Fortaleza: Sociedade Brasileira de Genética, Regional do Ceará, 2000. 11p. Site a ser disponibilizado.
- RESENDE, M.D.V. de. **Predição de valores genéticos, componentes de variância, delineamentos de cruzamento e estrutura de populações no melhoramento florestal**. Curitiba: Universidade Federal do Paraná, 1999. 434p. Tese Doutorado.
- 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 aditivos e genotípicos em espécies frutíferas. **Revista Brasileira de Fruticultura**, v.22, n.1, p.44-52, 2000.
- RESENDE, M.D.V. de; STURION, J.A.; CARVALHO, A.P. de ; SIMEÃO, R.M.; FERNANDES, J.S.C. **Programa de melhoramento da erva-mate coordenado pela Embrapa- Resultados da avaliação genética de populações, progêneres, indivíduos e clones**. Colombo: Embrapa Florestas, 2000a. 45p. (Embrapa Florestas. Circular Técnica, 43). no prelo.
- RESENDE, M.D.V. de; FURLANI JÚNIOR, E.; MORAES, M.L.T.; FAZUOLI, L.C. Estimação de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. **Bragantia**, 2000b. no prelo.
- RESENDE, M.D.V. de; FERNANDES, J.S.C. Procedimento BLUP individual para delineamentos experimentais aplicados ao melhoramento florestal. **Revista de Matemática e Estatística**. São Paulo, v.17. p.89-107, 1999.

- RESENDE, M.D.V. de; FERNANDES, J.S.C.; SIMEÃO, R.M. BLUP individual multivariado em presença de interação genótipo x ambiente para delineamentos experimentais repetidos em vários ambientes. **Revista de Matemática e Estatística**, São Paulo, v.17. p.209-228, 1999.
- RESENDE, M.D.V. de; MORA, A.L.; HIGA, A.R.; PALUDZYSZYN FILHO, E. Efeito do tamanho amostral na estimativa da herdabilidade em espécies perenes. **Floresta**, v.28, n.1/2, 1998.
- RESENDE, M.D.V.; PRATES, D.F.; JESUS, A.; YAMADA, C.K. Estimação de componentes de variância e predição de valores genéticos pelo método da máxima verossimilhança restrita (REML) e melhor predição linear não viciada (BLUP) em *Pinus*. **Boletim de Pesquisa Florestal**, Colombo, v.32/33, p.18-45, 1996.
- RESENDE, M.D.V. de.; VENCOVSKY, R.; FERNANDES, J.S.C. Selection and genetic gains in populations of *Eucalyptus* with a mixed mating system. In: IUFRO CONFERENCE, 1995, Hobart. **Proceedings** ... Hobart: IUFRO, 1995. p.191-193.
- RESENDE, M.D.V. de; HIGA, A.R. Estimação de valores genéticos no melhoramento de *Eucalyptus*: seleção em um caráter com base em informações do indivíduo e seus parentes. **Boletim de Pesquisa Florestal**, Colombo, n.28/29, p.11-36, 1994b.
- RESENDE, M.D.V. de; HIGA, A.R. Maximização da eficiência da seleção em testes de progêneres de *Eucalyptus* através da utilização de todos os efeitos do modelo matemático. **Boletim de Pesquisa Florestal**, Colombo, n.28/29, p.37-55, 1994a.
- RESENDE, M.D.V. de; HIGA, A.R.; LAVORANTI, O.J. Predição de valores genéticos no melhoramento de *Eucalyptus* – Melhor Predição Linear. In: CONGRESSO FLORESTAL BRASILEIRO, 7., 1993, Curitiba. **Anais**. São Paulo: Sociedade Brasileira de Silvicultura, 1993. v.1, p.144-147.
- SCHALL, R. Estimation in generalized linear models with random effects. **Biometrika**, London, v.78, p.719-727, 1991.
- SEARLE, S.R.; CASELLA, G.; McCULLOCH, C.E. **Variance components**. New York. J. Wiley, 1992. 528p.
- STEEL, R.G.D.; TORRIE, J.H. **Principles and procedures of statistics**. 2.ed. New York: Mc Graw-Hill, 1980. 633p.
- VENCOVSKY, R. Herança quantitativa. In: PATERNIANI, E.; VIEGA, G.P., eds. **Melhoramento e Produção do Milho**. Campinas: Fundação Cargill, 1987. p.137-214.
- VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486p.