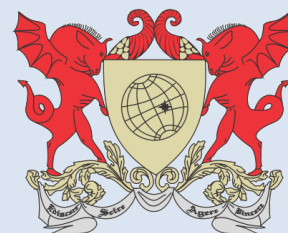


*Selegen Reml/Blup – Sistema Estatístico e Seleção Genética Computadorizada*  
*Manual Complementar do Selegen-Reml/Blup 2017*

( $M\alpha\tau\zeta\theta \int \partial\epsilon\theta\eta$  , Maio – 2017)



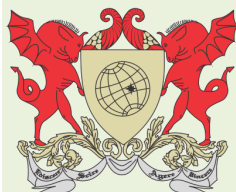
*Marcos Deon Vilela de Resende*



*Universidade Federal de Viçosa*  
*Centro de Ciências Exatas*  
*Departamento de Estatística*  
*Estatística Aplicada e Biometria*

**Viçosa – Minas Gerais**  
**2017**

## Sumário



1. Introdução [3]
2. Modelos com Matriz de Parentesco Completa (Modelos 165 a 169; 182) [4]
3. Modelos para Plantas Autógamas nas várias Gerações de Autofecundação (Modelos 170 a 179) [8]
4. Modelo para o BLUP Multigeracões no método Bulk dentro de progênies no Melhoramento de Plantas Autógamas (Modelo 180) [9]
5. Modelo Fatorial Interpopulacional para Linhagens com Interação G x E (Modelo 181) [12]
6. Comunalidade, Especificidade e Interação (Modelos 102 e 105) [14]
7. Modelo Linear Generalizado Misto para Variáveis Categóricas (Binomiais) [15]
8. Modelo para o BLUP com Genitores, Populações, Progênies e Gerações (SIPPPG) no método Bulk dentro de progênies no Melhoramento de Plantas Autógamas (Modelo 186) [20]
9. Avisos e outras Inclusões [21]
10. Novos modelos para Plantas Autógamas (Modelos 187 a 200) [23]
11. Referências Bibliográficas [28]

## Selegen Reml/Blup – Sistema Estatístico e Seleção Genética Computadorizada

(Ματζθ ∫ ∂εθη , Maio – 2015)

Marcos Deon Vilela de Resende



Universidade Federal de Viçosa  
Centro de Ciências Exatas  
Departamento de Estatística  
Estatística Aplicada e Biometria

### 1. Introdução

O software Selegen-Reml/Blup teve seu início em 1993 e possui agora mais de 22 anos. Software análogo é o Selegen Genômica que foi iniciado em 2007 e está agora em seu nono ano. Juntos cobrem a maioria das situações práticas encontradas por quem faz melhoramento genético de plantas perenes e anuais, englobando espécies alógamas, autógamas, de sistema reprodutivo misto e de reprodução assexuada (vegetativa e apomítica) (Resende et al., 1993 e 1996; Resende, 2000; 2002; 2007a e b; 2008). Algumas situações do melhoramento animal podem também ser executadas pelos dois softwares (Resende e Rosa-Perez, 1999; Resende et al., 2012; Resende et al., 2014).

De maneira geral, uma análise estatística completa envolve as seguintes quatro atividades: a estimação de componentes de média, a estimação de componentes de variância, a realização de testes de hipóteses e a inferência sobre a acurácia e precisão da estimação/predição. No contexto dos modelos mistos, a realização dessas atividades envolve a predição BLUP, a estimação REML, a análise de deviance, o cômputo da acurácia de predição e da variância do erro de predição, respectivamente. Os dois softwares (Selegen-Reml/Blup e Selegen Genômica) permitem todos esses procedimentos, mostrando os resultados nos arquivos .res e .fam. O Selegen Genômica propicia todos os resultados fornecidos pelo Selegen-Reml/Blup, porém usando coeficientes de parentesco obtidos com base em locos marcadores de DNA em vez de parentesco genealógico. Além disso, fornece estimativas dos efeitos genéticos de marcas e testa suas significâncias.

A qualidade da avaliação genotípica deve ser inferida preferencialmente com base na acurácia ( $\hat{r}_{gg}$ ). Em experimentos balanceados pode ser usada também a estatística F de Snedecor, conforme Resende e Duarte (2007). A relação entre F e acurácia é dada por:  $F = 1/(1 - \hat{r}_{gg}^2)$ . Na classe dos estimadores/preditores não

viciados, a precisão é dada pelo parâmetro variância do erro de predição (PEV). A estatística PEV é relacionada à acurácia por meio das equações:  $\hat{r}_{gg} = (1 - PEV / \sigma_g^2)^{1/2}$ ;  $PEV = (1 - \hat{r}_{gg}^2) \sigma_g^2$ , em que  $\sigma_g^2$  é a variação genotípica entre os materiais em avaliação. A acurácia é fornecida no arquivo .fam e a PEV pode ser obtida a partir da acurácia e da  $\sigma_g^2$ .

Os procedimentos estatísticos empregados na análise genética de dados experimentais devem ser capazes de considerar os seguintes fatores: desbalanceamento de dados, matriz de parentesco completa, procedimentos diferenciados para variáveis contínuas (modelos lineares mistos) e discretas (modelos lineares generalizados mistos). O software Selegen emprega esses procedimentos estatísticos adequados.

O manual completo do Selegen foi publicado em 2007 (Resende, 2007b). A presente publicação aborda alguns dos novos modelos que foram incluídos após 2007 e, portanto, complementa o livro mencionado. Atualmente estão disponíveis 200 modelos lineares mistos.

## **2. Modelos com Matriz de Parentesco Completa (Modelos 165 a 169)**

Esses modelos são adequados para quaisquer números de gerações. Os demais modelos, abordados no livro de 2007, usam até três gerações.

As seguintes instruções para os modelos 165 a 169 devem ser seguidas, com respeito à matriz de parentesco A completa:

- Indivíduos filhos com zeros não significativos nas observações (ou seja, indivíduos mortos ou sem observações) devem ser eliminados dos arquivos de pedigree e de dados;
- Indivíduos pais sem observações devem entrar no arquivo de pedigree e não devem entrar no arquivo de dados;
- Indivíduos pais com observações devem entrar no arquivo de pedigree e podem entrar no arquivo de dados. Nesse caso, os efeitos fixos (blocos, locais) não podem ser codificados como um diferente para cada genitor. Só podem ser usados dados com mais de um genitor por efeito fixo.

Encontra-se anexo um exemplo, com arquivo de pedigree e de dados (famílias de irmãos completos, incluindo os genitores), adequado para o modelo 166. A seguir são descritos os arquivos para cada modelo.

### **Modelo 169**

Para o caso de teste clonal com estrutura de família, instalado no delineamento de blocos incompletos com uma planta por parcela, o seguinte modelo pode ser analisado no software Selegen Reml/Blup:

$$y = Xr + Za + Wc + Sf + Tb + e,$$

, em que y é o vetor de dados, r é o vetor dos efeitos fixos (repetição, local) somados à média geral, a é o vetor dos efeitos genéticos aditivos individuais

(assumidos como aleatórios),  $c$  é o vetor dos efeitos de clone dentro de família de irmãos completos (aleatórios),  $f$  é o vetor dos efeitos de dominância de famílias de irmãos germanos ou capacidade específica de combinação (aleatórios),  $b$  é o vetor dos efeitos de blocos incompletos (aleatórios),  $e$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

Esse modelo pode ser encontrado na janela “Clones Aparentados / Matriz A Completa”, modelo 169 no lado esquerdo da tela do software Selegen Reml/Blup, na janela interna “Matriz A Completa”. Deve ser fornecido um arquivo de dados com a seguinte sequência de colunas: *Observação Indivíduo Repetição Clone Família Bloco Árvore Variáveis*. Deve ser fornecido também um arquivo de pedigree com a seguinte sequência de colunas: *Indivíduo Pai Mãe*. Todos os indivíduos devem constar na primeira coluna do arquivo de pedigree, mesmo os que são apenas genitores. Genitores de genitores devem receber códigos zero na segunda e terceira colunas.

Os componentes de variância associados ao modelo, conforme notação do Selegen-Reml/Blup, são assim interpretados:

$V_a$ : variância genética aditiva.

$V_{clone/fam}$ : variância entre clones dentro de famílias de irmãos completos, ajustada para variância genética aditiva total.

$V_{fam}$ : variância da capacidade específica de combinação ou variância genética de dominância entre famílias de irmãos germanos.

$V_{bloc}$ : variância entre blocos.

$V_e$ : variância ambiental.

$V_f$ : variância fenotípica individual.

$h_{2a} = h_2$ : herdabilidade individual no sentido restrito no bloco, ou seja, dos efeitos aditivos.

$h_{2g}$ : herdabilidade individual no sentido amplo, ou seja, dos efeitos genotípicos totais.

$c_{2clone/fam} = c_2$ : coeficiente de determinação dos efeitos de clones dentro de famílias de irmãos completos, ajustado para variância genética aditiva total.

$c_{2fam} = c_{21}$ : coeficiente de determinação dos efeitos da capacidade específica de combinação.

$c_{2bloc} = c_{22}$ : coeficiente de determinação dos efeitos de bloco.

$u$  = Média geral do experimento.

Os componentes de variância apresentados acima podem ser decompostos da seguinte forma:

$V_a = \sigma_a^2$ : variância genética aditiva.

$V_{fam} = (1/4)\sigma_d^2 + (1/4)\sigma_{aa}^2 + (1/8)\sigma_{ad}^2 + (1/16)\sigma_{dd}^2 \dots$

$V_{clone/fam} = (3/4)\sigma_d^2 + (3/4)\sigma_{aa}^2 + (7/8)\sigma_{ad}^2 + (15/16)\sigma_{dd}^2 \dots$ , em que  $d$  refere-se a efeitos de dominância e  $aa$ ,  $ad$  e  $dd$  referem-se a efeitos epistáticos.

A herdabilidade da segregação mendeliana desregressada dos dados corrigidos  $y_c$  é dada por: 
$$h_m^2 = \frac{n \cdot 0.5h^2}{n \cdot 0.5h^2 + 1 - h^2 - c_{clone/fam}^2 - c_{fam}^2 - c_{bloc}^2}$$
, em que  $n$  é o número de repetições do clone. Expressa de outra forma, tem-se: 
$$h_m^2 = \frac{n \cdot 0.5\sigma_g^2}{n \cdot 0.5\sigma_g^2 + \sigma_e^2}.$$

Sob esse modelo o valor genotípico dos clones é dado por  $VG = \hat{u} + \hat{a} + \hat{c} + \hat{f} = \hat{u} + \hat{a} + \hat{c}_{clone/fam} + \hat{f}_{fam}$ .

#### Modelo 168 (3 efeitos aleatórios adicionais)

O seguinte modelo é analisado no software Selegen Reml/Blup:

$y = Xr + Za + Wv_1 + Sv_2 + Tv_3 + e$ , em que  $v_1$ ,  $v_2$  e  $v_3$  são efeitos aleatórios adicionais. Deve ser fornecido um arquivo de dados com a seguinte seqüência de colunas: *Observação Indivíduo Repetição EfeitoAdicional1 EfeitoAdicional2 EfeitoAdicional3 Árvore Variáveis*.

#### Modelo 167 (2 efeitos aleatórios adicionais)

O seguinte modelo é analisado no software Selegen Reml/Blup:

$y = Xr + Za + Wv_1 + Sv_2 + e$ , em que  $v_1$  e  $v_2$  são efeitos aleatórios adicionais. Deve ser fornecido um arquivo de dados com a seguinte seqüência de colunas: *Observação Indivíduo Repetição EfeitoAdicional1 EfeitoAdicional2 Árvore Variáveis*.

#### Modelo 166 (1 efeito aleatório adicional)

O seguinte modelo é analisado no software Selegen Reml/Blup:

$y = Xr + Za + Wv_1 + e$ , em que  $v_1$  é o efeito aleatório adicional. Deve ser fornecido um arquivo de dados com a seguinte seqüência de colunas: *Observação Indivíduo Repetição EfeitoAdicional1 Árvore Variáveis*.

#### Modelo 165 (sem efeito aleatório adicional)

O seguinte modelo é analisado no software Selegen Reml/Blup:

$y = Xr + Za + e$ . Deve ser fornecido um arquivo de dados com a seguinte seqüência de colunas: *Observação Indivíduo Repetição Árvore Variáveis*.

#### Matriz de parentesco igual à identidade

Os modelos 165 a 169 podem também ser usados para testes clonais e de tratamentos (híbridos, famílias) não aparentados. Nesse caso, o arquivo de pedigree deve conter a numeração dos tratamentos na primeira coluna e zero nas colunas de pai e mãe.

#### Modelo 182 (efeitos aleatórios de família, clone e bloco)

Este modelo 182 é adequado para uma situação similar à do modelo 169, em que se dispõe de um teste clonal com clones provenientes de várias famílias de irmãos completos (também chamado testes de progênies clonadas). É uma situação

comum no melhoramento da cana-de-açúcar, da mandioca e do eucalipto. Entretanto, difere do modelo 169 pelo fato de não considerar a matriz de parentesco completa. E por isso o modelo 182 é mais rápido. É análogo também ao modelo 162, que é adequado a cruzamentos dialélicos. Mas o modelo 182 é também mais rápido e simples (e tão eficiente quanto) que o modelo 162.

Para o caso de teste clonal com estrutura de família, instalado no delineamento de blocos incompletos com uma planta por parcela, o seguinte modelo é ajustado:

$$y = Xr + Zf + Wb + Sc + e,$$

, em que  $y$  é o vetor de dados,  $r$  é o vetor dos efeitos fixos (repetição, local) somados à média geral,  $f$  é o vetor dos efeitos de famílias de irmãos germanos (aleatórios),  $b$  é o vetor dos efeitos de blocos incompletos (aleatórios),  $c$  é o vetor dos efeitos de clone dentro de família de irmãos completos (aleatórios),  $e$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

Esse modelo pode ser encontrado na janela “Clones Aparentados / Matriz A Completa”, modelo 182 no lado esquerdo da tela do software Selegen Reml/Blup, na janela interna “Blocos Incompletos”. Deve ser fornecido um arquivo de dados com a seguinte sequência de colunas: *Observação Família Repetição Bloco Clone Árvore Variáveis*.

Os componentes de variância associados ao modelo, conforme notação do Selegen-Reml/Blup, são assim interpretados:

Vfam: variância total entre famílias de irmãos germanos.

Vbloc: variância entre blocos.

Vclone/fam: variância entre clones dentro de famílias de irmãos completos, **não** ajustada para variância genética aditiva total.

Ve: variância ambiental.

Vf: variância fenotípica individual.

h2fam = h2: coeficiente de determinação dos efeitos de famílias de irmãos completos.

h2g: herdabilidade individual no sentido amplo, ou seja, dos efeitos genotípicos totais.

c2clone/fam = c2: coeficiente de determinação dos efeitos de clones dentro de, **não** ajustado para variância genética aditiva total.

c2bloc = c22: coeficiente de determinação dos efeitos de bloco.

u = Média geral do experimento.

Os componentes de variância apresentados acima podem ser decompostos da seguinte forma:

$Va = \sigma_a^2$ : variância genética aditiva.

$Vfam = (1/2)\sigma_a^2 + (1/4)\sigma_d^2$

$Vclone/fam = (1/2)\sigma_a^2 + (3/4)\sigma_d^2 + \sigma_{ie}^2$ , em que  $ie$  referem-se aos efeitos de interação epistática.

A herdabilidade total é dada por:  $h_g^2 = \frac{V_{fam} + V_{clone}/fam}{V_f}$ .

Sob esse modelo o valor genotípico dos clones é dado por  $VG = \hat{u} + \hat{f} + \hat{c} = \hat{u} + \hat{f}_{fam} + \hat{c}_{clone}/fam$ . Assim, o segundo ranking (clone + família) mostrado no arquivo .res deve ser o utilizado para a seleção de clones.

O modelo é igualmente válido para famílias de meios irmãos e irmãos S1. Nesse caso tem-se as seguintes interpretações para os componentes de variância associados ao modelo:

#### **Irmãos S1**

$$V_{fam} = \sigma_a^2 + (1/4)\sigma_d^2$$

$$V_{clone}/fam = (1/2)\sigma_a^2 + (1/2)\sigma_d^2 + \sigma_{ie}^2$$

#### **Meios Irmãos**

$$V_{fam} = (1/4)\sigma_a^2$$

$$V_{clone}/fam = (3/4)\sigma_a^2 + \sigma_d^2 + \sigma_{ie}^2$$

#### **Modelo 184 (Matriz de parentesco genético aditivo A e sua inversa A<sup>-1</sup>)**

Esse modelo cria e grava a matriz de parentesco genético aditivo A e sua inversa A<sup>-1</sup>. A matriz A pode ser usada diretamente no modelo 9 do software Selegen Genômica.

### **3. Modelos para Plantas Autógamas nas várias Gerações de Autofecundação (Modelos 170 a 179)**

Foram desenvolvidos 10 modelos (modelos 170 a 179) no Selegen-Reml/Blup e incorporados na janela “Espécies Perenes Autógamas: Café Arábica, Pêssego”. Os modelos envolvem as gerações S0 (irmãos germanos) a S5 com o uso de medidas repetidas (modelos 170 a 175) e gerações S2 a S5 sem o uso de medidas repetidas (modelos 176 a 179).

Os modelos 170 a 175 ajustam os efeitos genéticos aditivos, de parcela, permanentes de indivíduos, da interação famílias x colheitas e de blocos incompletos. Os modelos 176 a 179 logicamente não ajustam os efeitos permanentes de indivíduos e da interação famílias x colheitas. Quando não se tem blocos incompletos e/ou parcelas, esses devem ter o c24 e/ou o c2 zerados na tela do Selegen.

Os seguintes parâmetros são estimados pelos modelos 170 a 175:

Vat: variância aditiva total.

Vparc: variância entre parcelas.

Vperm: variância dos efeitos permanentes.



Vint: variância da interação genótipos x colheitas.  
 Vbloc: variância entre blocos.  
 Ve: variância residual.  
 Vf: variância fenotípica  
 h2a: herdabilidade individual no sentido restrito, com somente a variação entre progênie no numerador.  
 h2ad: herdabilidade aditiva individual dentro de progênie.  
 h2atc: herdabilidade individual total no sentido restrito, ajustada para os efeitos do modelo.  
 czparc: coeficiente de determinação dos efeitos de parcela.  
 czperm: coeficiente de determinação dos efeitos permanente.  
 czint: coeficiente de determinação dos efeitos da interação progênie x colheitas.  
 czbloc: coeficiente de determinação dos efeitos de bloco.  
 rgmed: correlação genotípica entre o desempenho das progênie nas duas safras.  
 r: coeficiente de repetibilidade individual.

Nos modelos 176 a 179, h2a refere-se à herdabilidade individual no sentido restrito, com somente a variação entre progênie no numerador.

Uma listagem dos modelos é apresentada a seguir.

- 170. Seleção em S0 (Irmãos Germanos) com Medidas Repetidas
- 171. Seleção em S1 com Medidas Repetidas
- 172. Seleção em S2 com Medidas Repetidas
- 173. Seleção em S3 com Medidas Repetidas
- 174. Seleção em S4 com Medidas Repetidas
- 175. Seleção em S5 com Medidas Repetidas
- 176. Seleção em S2 sem Medidas Repetidas
- 177. Seleção em S3 sem Medidas Repetidas
- 178. Seleção em S4 sem Medidas Repetidas
- 179. Seleção em S5 sem Medidas Repetidas

Para os modelos 170 a 175 devem ser fornecidos arquivos de dados com a seguinte sequência de colunas: *Observação Família Medição Parcela Indivíduo Interação Bloco Árvore Variáveis*. Para os modelos 176 a 179 devem ser fornecidos arquivos de dados com a seguinte sequência de colunas: *Observação Família Repetição Parcela Bloco Árvore Variáveis*.

#### **4. Modelo para o BLUP Multigerações no método Bulk dentro de progênie no Melhoramento de Plantas Autógamas**

O método BLUP no Bulk Dentro de Progênie (BDP-BLUP) foi proposto para o melhoramento de plantas autógamas por Resende, Ramalho et al. (2014). É muito útil para espécies como o feijão, a soja, o arroz e o trigo. O modelo 180, na janela de Plantas Autógamas, contempla esse método. Tal método, para as gerações F3 ou S1 até F6 ou S4, baseia-se nos seguintes pesos, dados aos efeitos ajustados (via Blup simultâneo) de progênie nas 4 gerações (S1 a S4):

$$\begin{bmatrix} (K_1/h^2) & 1.00 & 1.00 & 1.00 \\ & (K_2/h^2) & 1.50 & 1.50 \\ Sim. & & (K_3/h^2) & 1.75 \\ & & & (K_4/h^2) \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \end{bmatrix} = \begin{bmatrix} 1.00 \\ 1.50 \\ 1.75 \\ 1.875 \end{bmatrix} \text{ em que:}$$

$$K_i = \frac{(1 + Fh^2) + (n-1)rh^2}{n}, \text{ sendo } F \text{ o coeficiente de endogamia dos indivíduos de cada}$$

geração;

$r$  é a correlação genética aditiva entre indivíduos de cada progênie que equivalem a 1 para progênies  $S_1$  ou  $F_3$ , 1,5 para  $S_2$  ou  $F_4$ , 1,75 para  $S_3$  ou  $F_5$  e 1,875 para  $S_4$  ou  $F_6$ .

Assim, sendo  $n_{Si}$  o número de repetições em cada geração, tem-se:

$$K_1 = \frac{(1 + 0.5h^2) + (n_{S1} - 1)h^2}{n_{S1}}$$

$$K_2 = \frac{(1 + 0.75h^2) + (n_{S2} - 1)1.5h^2}{n_{S2}}$$

$$K_3 = \frac{(1 + 0.875h^2) + (n_{S3} - 1)1.75h^2}{n_{S3}}$$

$$K_4 = \frac{(1 + 0.9375h^2) + (n_{S4} - 1)1.875h^2}{n_{S4}}$$

O modelo ajustado é o seguinte:  $y = Xg + Za + Wb + Ti + e$ , em que  $y$  é o vetor de dados,  $g$  é o vetor dos efeitos de geração (assumidos como fixos) somados à média geral,  $a$  é o vetor dos efeitos genéticos aditivos de famílias (assumidos como aleatórios),  $b$  é o vetor dos efeitos de blocos (assumidos como aleatórios),  $i$  é vetor dos efeitos da interação famílias x gerações (aleatórios) e  $e$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos. A seguinte sequência de colunas é exigida no arquivo de dados:

*Geração Parcela Família Geração Bloco Interação Nrep Variáveis*

A primeira coluna, referente à gerações, deve ser codificada com  $S_1$ ,  $S_2$ ,  $S_3$  e  $S_4$ . Se uma ou algumas gerações não estiverem presentes, não há problema. O BLUP será computado com as gerações existentes, ou seja, no mínimo uma e no máximo quatro. Gerações de autofecundação além de  $S_4$  ou  $F_6$  não podem ser utilizadas, ou se usadas será de forma aproximada e deverão ser codificadas como  $S_4$ .

A coluna Interação deve codificar combinações geração-famílias. Assim, o código 12 representa geração 1, família 2. No Excel, essa codificação pode ser obtida pelo comando  $A2 \& C2$ , quando se tem a codificação de gerações na coluna A e de famílias na coluna C. Outras codificações similares são também válidas.

A numeração dos blocos deve ser sequencial de uma repetição para outra e de uma geração para outra, ou seja, os números dados aos blocos devem ser diferentes em cada repetição e geração.

A coluna Nrep deve conter o número de repetições usadas na experimentação em cada geração. Assim, deve conter no máximo 4 Algarismos diferentes, um para  $S_1$ ,  $S_2$ ,  $S_3$  e  $S_4$ . Se o mesmo número de repetições for adotado, ou seja, for constante através das gerações, a coluna será toda preenchida com o mesmo algarismo. Sob um modelo puramente aditivo, os seguintes componentes de variância são estimados:

Va: variância genética aditiva.

Vbloc: variância ambiental entre blocos.

Vint: variância da interação famílias x gerações.

Ve: variância residual.

Vf: variância fenotípica individual.

h<sub>2a</sub> = h<sub>2</sub>: herdabilidade de parcelas individuais no sentido restrito, ou seja, dos efeitos aditivos.

c<sub>2b</sub> = c<sub>2</sub>: coeficiente de determinação dos efeitos de bloco.

c<sub>2int</sub> = c<sub>2i</sub>: coeficiente de determinação dos efeitos da interação famílias x gerações.

rgger: correlação genotípica entre o desempenho nas várias gerações.

CV<sub>gi</sub>‰: coeficiente de variação genotípica.

CV<sub>e</sub>‰: coeficiente de variação residual.

Média geral do experimento.

Os seguintes valores genotípicos são estimados:

*Seleção de famílias – comuns a todas as gerações (livre da interação com gerações)*

Ordem	Família	g	u+g	Ganho	Nova Média	u+g+gem
1	9	0.0385	0.2179	0.0385	0.2179	0.2402
2	7	0.0316	0.2111	0.0351	0.2145	0.2327
3	18	0.0191	0.1985	0.0297	0.2092	0.2188
4	1	0.0164	0.1959	0.0264	0.2058	0.2159
5	19	0.0128	0.1923	0.0237	0.2031	0.2119

Esses resultados são referentes aos efeitos (g) e valores (u + g) genéticos aditivos preditos, livres de toda interação com gerações. A quantidade (u + g + gem) refere-se ao valor genotípico médio nas várias gerações e capitaliza uma interação média com todas as gerações avaliadas.

*Seleção de famílias por geração (com interação com gerações)*

Geração	Ordem	Família	g+ge	u+g+ge	Ganho	Nova Média
1	1	9	0.0485	0.2508	0.0485	0.2508
	2	7	0.0365	0.2387	0.0425	0.2447
	3	2	0.0250	0.2272	0.0367	0.2389
	4	18	0.0207	0.2229	0.0327	0.2349
	5	1	0.0185	0.2207	0.0298	0.2321
2	1	15	0.0056	0.1839	0.0056	0.1839
	2	8	0.0055	0.1838	0.0056	0.1838
	3	7	0.0051	0.1834	0.0054	0.1837
	4	3	0.0051	0.1833	0.0053	0.1836
	5	18	0.0044	0.1826	0.0051	0.182634

Nessa seção são apresentados os valores genotípicos preditos para cada família em cada geração ( $u + g + ge$ ), ou seja, os valores genotípicos capitalizando a interação com os gerações.

### Valores Genéticos Finais na Geração $S_{\infty}$ ( $VG_{\infty}$ )

Estes são obtidos ponderando-se os resultados acima ( $g + ge$ ), de famílias por geração, pelos pesos  $b_1$  a  $b_4$  e depois somando-se a média geral. Para cômputo dos pesos  $b_1$  a  $b_4$  são usados a  $h^2$  individual do caráter (estimada por REML) e os números  $n$  de repetições em cada geração, informados na coluna Nrep. Resultados idênticos podem ser obtidos por um BLUP multivariado.

#### Seleção Final de Famílias (BDP-BLUP)

Ordem	Genótipo	$VG_{\infty}$
1	5	0.2176
2	9	0.2098
3	7	0.2074
4	18	0.2008
5	1	0.1991

A seleção final deve basear-se nos valores genotípicos preditos na geração  $S_{\infty}$  ( $VG_{\infty}$ ).

Esse modelo 180 pode ser usado com qualquer número de gerações de  $F_3$  a  $F_6$ . As seguintes situações são contempladas:

Gerações	Pesos
S1 S2 S3 S4	b1 b2 b3 b4
S1 S2 S3	b1 b2 b3
S1 S2 S4	b1 b2 b4
S1 S3 S4	b1 b3 b4
S2 S3 S4	b2 b3 b4
S1 S2	b1 b2
S1 S3	b1 b3
S1 S4	b1 b4
S2 S3	b2 b3
S2 S4	b2 b4
S3 S4	b3 b4
S1	b1
S2	b2
S3	b3
S4	b4

## 5. Modelo Fatorial Interpopulacional para Linhagens com Interação $G \times E$

Este modelo (181) é usado no caso de cruzamentos entre linhagens completamente endogâmicas. É muito utilizado por empresas privadas no melhoramento do milho. Os componentes  $V_{gm}$  e  $V_{gf}$  estimam  $(1/2)$  da variação genética aditiva e o componente  $V_{cec}$  estimará toda a variação genética de dominância.

O modelo estatístico é dado por  $y = Xr + Zm + Wf + Tc + Si + Qb + e$ , em que  $y$  é o vetor de dados,  $r$  é o vetor dos efeitos de repetição (assumidos como fixos)

somados à média geral,  $m$  é o vetor dos efeitos de genitores da população 1 ou população de machos (não necessariamente) (assumidos como aleatórios),  $f$  é o vetor dos efeitos genitores da população 2 ou população de fêmeas (não necessariamente) (aleatórios),  $c$  é o vetor dos efeitos da capacidade específica de combinação dos genitores da população 1 com os genitores da população 2 (aleatórios),  $i$  é o vetor dos efeitos da interação progênie com locais (aleatórios)  $i_c$  é o vetor dos efeitos da interação da cec com locais (aleatórios),  $b$  é o vetor dos efeitos de bloco (aleatórios),  $e$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

O modelo 181 situa-se na janela “Fatoriais Interpopulacionais”, no lado direito da tela do software Selegen Reml/Blup, na janela interna “Blocos Incompletos”. A seguinte seqüência de colunas é exigida no arquivo de dados:  
*Indivíduo Machos Repetição Fêmeas Progênie Interação Bloco Árvore Variáveis*

Os seguintes componentes de variância são estimados:

Vgm: variância genética entre machos ou genitores da população 1 em cruzamento com a população 2, a qual estima (1/2) da variação genética aditiva correspondente.

Vgf: variância genética entre fêmeas ou genitores da população 2 em cruzamento com a população 1, a qual estima (1/2) da variação genética aditiva correspondente.

Vcec: variância da capacidade específica de combinação interpopulacional entre dois genitores, a qual estima toda a variação genética de dominância correspondente.

Vint: variância do efeito da interação progênie x locais, contemplando a interação dos efeitos de híbridos (aditivo total + dominante total) com ambientes.

Vbloc: variância entre blocos.

Ve: variância residual.

Vf: variância fenotípica individual.

$c_{2m} = h_2$ : coeficiente de determinação dos efeitos de genitores na população 1, ou seja, dos efeitos aditivos interpopulacionais.

$c_{2f} = c_2$ : coeficiente de determinação dos efeitos de genitores na população 2, ou seja, dos efeitos aditivos interpopulacionais.

$c_{2cec} = c_{21}$ : coeficiente de determinação dos efeitos da capacidade específica de combinação.

$c_{2int} = c_{24}$ : coeficiente de determinação dos efeitos da interação progênie x locais.

$c_{2bloc} = c_{22}$ : coeficiente de determinação dos efeitos de bloco.

$h_{2am}$ : herdabilidade interpopulacional individual no sentido restrito na população 1.

$h_{2af}$ : herdabilidade interpopulacional individual no sentido restrito na população 2.

h2a: herdabilidade interpopulacional individual no sentido restrito, média para as duas populações.

h2dom: herdabilidade individual dos efeitos interpopulacionais de dominância.

h2g: herdabilidade interpopulacional individual no sentido amplo, ou seja, dos efeitos genotípicos totais.

Média geral do experimento.

## 6. Comunalidade, Especificidade e Interação (Modelos 102 e 105)

O conceito de **comunalidade** entre duas variáveis é o mesmo que o de uma correlação proporcional a uma soma de correlações. Seu numerador é a **variância comum** ou covariância e seu denominador é função da variância comum e das **variâncias específicas**. O conceito de **especificidade** é relacionado ao conceito de **interação** entre duas variáveis. Estimativas de variância da interação, especificidade média, comunalidade e proporção especificidade média/comunalidade foram incluídas nos modelos 102 e 105. Um estimador para a variância da interação entre duas variáveis é derivado a seguir.

Tomando-se a variância das diferenças entre duas variáveis  $x_1$  e  $x_2$  tem-se:

$Var(x_1 - x_2) = Var(x_1) + Var(x_2) - 2Cov(x_1, x_2)$ . Sendo  $Cov(x_1, x_2)$  a variância comum ( $Var(x)$ ) explicada por ambas, tem-se:  $Var(x_1 - x_2) = Var(x_1) + Var(x_2) - 2Var(x)$ . Dividindo por 2 ambos os lados da equação, tem-se:

$$\frac{Var(x_1 - x_2)}{2} = \frac{Var(x_1) + Var(x_2)}{2} - Var(x).$$

Encontra-se demonstrado na literatura que  $\frac{Var(x_1) + Var(x_2)}{2} = Var(x) + Var_{int}(x_1, x_2)$ , em que  $Var_{int}(x_1, x_2)$  é a variância da interação entre as duas variáveis. Dessa expressão tem-se que  $Var_{int}(x_1, x_2) = \frac{Var(x_1) + Var(x_2)}{2} - Var(x)$ . Comparando-se essa expressão com a expressão para a variância das diferenças entre duas variáveis, verifica-se que  $Var_{int}(x_1, x_2) = \frac{Var(x_1 - x_2)}{2}$  e, portanto, a **variância da interação entre duas variáveis** pode ser estimada como a **metade da variância das diferenças entre as duas variáveis**. Dessa forma, não há necessidade de análise de variância conjunta das variáveis para a estimação da variância da interação.

Outra forma de estimar a variância da interação é por meio de  $Var_{int}(x_1, x_2) = \frac{Var(x_1) + Var(x_2)}{2} - Var(x)$ , a qual é equivalente a  $Var_{int}(x_1, x_2) = \frac{Var(x_1) + Var(x_2)}{2} - Cov(x_1, x_2)$ . Assim,  $Var_{int}(x_1, x_2)$  é estimada como a **média das variâncias das duas variáveis menos a covariância entre as duas variáveis**. Verifica-se então que o uso de informações correlacionadas pode ser realizado via inclusão de um fator de efeitos da interação.

Como exemplo, considere:

$$Var(x_1) = 0.7448; \quad Var(x_2) = 0.1904; \quad Cov(x_1, x_2) = 0.3074; \quad Var(x_1 - x_2) = 0.3204.$$

As seguintes quantidades podem ser computadas:

**Correlação:**  $Cor(x_1, x_2) = \frac{Cov(x_1, x_2)}{\sqrt{Var(x_1)Var(x_2)}} = \frac{0.3074}{\sqrt{0.7448 \cdot 0.1904}} = 0.816$

**Variância da interação :**  $Var_{int}(x_1, x_2) = \frac{Var(x_1 - x_2)}{2} = \frac{0.3204}{2} = 0.1602$

**Variância da interação :**  $Var_{int}(x_1, x_2) = \frac{Var(x_1) + Var(x_2)}{2} - Cov(x_1, x_2) = \frac{0.7448 + 0.1904}{2} - 0.3074 = 0.1602$

**Especificidade Média ou Variância da interação padronizada :**

**Especificidade Média**  $S_M(x_1, x_2) = \frac{Var_{int}(x_1, x_2)}{\sqrt{Var(x_1)Var(x_2)}} = \frac{0.1602}{\sqrt{0.7448 \cdot 0.1904}} = \frac{Var_{int}(x_1, x_2) \cdot Cor(x_1, x_2)}{Cov(x_1, x_2)} = \frac{0.1602 \cdot 0.816}{0.3074} = 0.4253$

**Comunalidade:**

**Comunalidade**  $Com(x_1, x_2) = \frac{Cor(x_1, x_2)}{Cor(x_1, x_2) + S_M(x_1, x_2)} = \frac{0.816}{0.816 + 0.4253} = \frac{Cov(x_1, x_2)}{Cov(x_1, x_2) + Var_{int}(x_1, x_2)} = \frac{0.3074}{0.3074 + 0.1602} = 0.6574$

A variância da interação equivale a uma média das variâncias específicas de cada variável. A comunalidade apresenta menor valor do que a correlação e, portanto, é um estimador mais conservador de associação do que o é a correlação. Quanto maior a comunalidade menor a interação e vice-versa. Essas informações são muito importantes nos estudos de interação genótipos x locais e genótipos x medições, informando sobre as correlações genéticas e interações entre ambientes, dois a dois. Para isso, os diferentes ambientes (locais ou colheitas) devem ser colocados como variáveis distintas (diferentes colunas) no Selegen. Outra quantidade importante estimada é a proporção variância da interação / covariância que é dada por  $\frac{Var_{int}(x_1, x_2)}{Cov(x_1, x_2)} = \frac{0.1602}{0.3074} = \frac{1 - Com(x_1, x_2)}{Com(x_1, x_2)} = \frac{1 - 0.6574}{0.6574} = 0.5212$

A correlação ordinária e a comunalidade são dois conceitos alternativos de **congruência** ou covariação padronizada ou escalada, ou seja, de correlação. Na correlação, a covariância é escalada (dividida) pela **média geométrica das variâncias**, ou seja,  $Cor(x_1, x_2) = \frac{Cov(x_1, x_2)}{\sqrt{Var(x_1)Var(x_2)}}$ . Na comunalidade, a covariância é escalada (dividida) pela

**média aritmética das variâncias**, ou seja,  $Com(x_1, x_2) = \frac{Cov(x_1, x_2)}{Cov(x_1, x_2) + Var_{int}(x_1, x_2)} = \frac{Cov(x_1, x_2)}{\frac{Var(x_1) + Var(x_2)}{2}}$ . A inter-

relação entre correlação, comunalidade e variância específica média ou variância da interação é dada por  $Cor(x_1, x_2) = Com(x_1, x_2) + Var_{int}(x_1, x_2) = 0.6574 + 0.1602 = 0.817$ . Portanto, tem-se também  $Com(x_1, x_2) = Cor(x_1, x_2) - Var_{int}(x_1, x_2)$ , ou seja, a diferença entre correlação e variância da interação fornece a comunalidade. Assim, esta equivale a uma correlação livre da interação e é então de grande importância ao melhoramento.

Em análise multivariada, a comunalidade é usada em associação com a técnica da análise de fatores. Neste contexto, assume um conceito análogo ao da confiabilidade no caso univariado.

## 7. Modelo Linear Generalizado Misto para Variáveis Categóricas (Binomiais)

Para variáveis binomiais, avaliadas por 0 e 1, foi desenvolvido o modelo 183 (incluído na janela “Dados Categóricos”), seguindo o desenvolvimento teórico apresentado por Resende (2002). A função de ligação empregada é a logito. A composição dos arquivos de dados e de pedigree segue a forma especificada para o modelo 166.



Esse modelo pode também ser usado para testes clonais e de tratamentos (híbridos, famílias) não aparentados. Nesse caso, o arquivo de pedigree deve conter a numeração dos tratamentos na primeira coluna e zero nas colunas de pai e mãe.

As seguintes abordagens podem ser usadas para estimar modelos com variáveis dicotômicas: **modelo de probabilidade linear**; **modelo logit**; **modelo probit** (ou **normit**). De maneira geral, esses dois últimos são mais adequados.

Os modelos probit e logit são bem comparáveis e a principal diferença entre eles refere-se ao fato de a distribuição logística ter caldas ligeiramente mais achatadas, isto é, a curva normal ou probit se aproxima do eixo mais rapidamente do que a curva logística. A distribuição logística é muito similar à distribuição t com sete graus de liberdade e a distribuição normal é uma distribuição t com infinitos graus de liberdade.

Um modelo linear generalizado é definido por: (i) um componente aleatório associado à distribuição da variável resposta (dados observados); (ii) um componente sistemático linear nos parâmetros, denominado preditor linear ou estrutura linear do modelo; (iii) uma função de ligação, a qual combina o componente aleatório e o componente sistemático. A **componente aleatória** identifica a família de distribuições da variável resposta, a **componente sistemática** constitui uma função linear de variáveis explicativas que atua como um preditor e a **função de ligação** conecta a componente sistemática a uma função da média das observações.

No caso da variável sobrevivência, a probabilidade  $\mu$ , de uma planta sobreviver (receber o escore 1) é função decrescente da dosagem do ambiente adverso, obedecendo à função sigmóide dada por:  $P(Y=1) = \mu = \frac{1}{1+e^{-\theta}} = \frac{e^{\theta}}{1+e^{\theta}}$ , em que  $\theta$  é a variável na escala latente logística, descrita por um modelo linear. Esta função é denominada função distribuição acumulada logística.

Assim, o logito de  $\mu$  equivale a  $\eta = \theta$ , e um modelo linear pode ser imposto a  $\eta$ . Substituindo-se os valores obtidos de  $\theta$  na função sigmóide [ $\mu = e^{\theta} / (1 + e^{\theta})$ ], produz-se  $\mu$  no intervalo entre 0 e 1.

Característica do modelo logit: conforme  $\mu_i$  vai de 0 a 1 (isto é, conforme  $\theta_i$  varia de  $-\infty$  a  $+\infty$ ), o logit vai de  $-\infty$  a  $+\infty$ , ou seja, embora as probabilidades  $\mu_i$  se situem entre 0 e 1 os logits não se restringem a este limite; embora o logit seja linear na variável observada, as probabilidades  $\mu_i$  não são. Essa propriedade contrasta com o modelo linear em que as probabilidades  $\mu_i$  aumentam linearmente com a variável observável.

Em resumo, na definição de um MLG, são fundamentais: (i) definição da distribuição da variável resposta; (ii) definição do preditor linear ou matriz do modelo; (iii) definição da função de ligação.

Considerando a variável sobrevivência, em que os indivíduos vivos recebem o valor 1 e os mortos o valor 0, tem-se que:



$\mu_i = \text{Prob} (Y_i = 1)$  : probabilidade de o indivíduo sobreviver;

$g(\mu_i) = \eta_i = \log \left[ \frac{\mu_i}{1 - \mu_i} \right]$ : logito da observação  $i$  ou função de ligação que liga a tolerância

(escala latente, função logística) a média  $\mu_i$  do dado binário  $y_i$ ;

$\eta = X\beta + Za$  : modelo linear imposto ao logito, em que:

$\beta$  e  $a$  : parâmetros de locação na escala logito (escala de tolerância).

$X$  e  $Z$  : matrizes de incidência para  $\beta$  e  $a$ , respectivamente.

Não existe termo de erro em  $\eta$ , diferentemente dos modelos lineares. A variância residual, ou seja, a variância das observações, dado  $\eta_i$  (ou  $\mu_i$ ), advém da amostragem binomial, ou seja:

$$\text{Var} (Y_i | \mu_i) = \mu_i (1 - \mu_i).$$

No caso, a variável de trabalho é definida pelo modelo  $y^* = X\beta + Za + (y - \mu) g'(\mu)$

$y^* = \eta + D^{-1} (y - \mu)$ : variável de trabalho, a qual é uma combinação linear do preditor  $\eta$  e da discrepância entre os valores observados e ajustados.

$D^{-1}$  : uma matriz diagonal com elementos  $(\partial \eta_i / \partial \mu_i) = \frac{1}{\mu_i (1 - \mu_i) \sigma_{e_L}^2}$ , que são os

pesos da análise.

$$\mu_i = \frac{e^{\eta_i}}{(1 + e^{\eta_i})}.$$

$$\text{Assim, } y_i^* = \eta_i + \frac{y_i - \mu_i}{\mu_i (1 - \mu_i)} \text{ e } \eta = \log \left( \frac{\mu}{1 - \mu} \right).$$

Assim tem-se:

$$\begin{aligned} y_i^* &= \eta_i + \frac{y_i - \mu_i}{\mu_i (1 - \mu_i)} \\ &= \log \left( \frac{\mu_i}{1 - \mu_i} \right) + \frac{y_i - \mu_i}{\mu_i (1 - \mu_i)}. \end{aligned}$$

Assim, têm-se as seguintes equações de modelo misto para o BLUP:

$$\begin{bmatrix} X' S^{-1} X & X' S^{-1} Z \\ Z' S^{-1} X & Z' S^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta}_L \\ \hat{a}_L \end{bmatrix} = \begin{bmatrix} X' S^{-1} y^* \\ Z' S^{-1} y^* \end{bmatrix}, \text{ em que:}$$

$S^{-1} = W' D^{-1} W$ : 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*);

$\beta_L$  e  $a_L$  = efeitos fixos e aleatórios na escala logística;

$W$  : uma matriz diagonal com elementos  $w_i = [\mu_i (1 - \mu_i)]$ , que são os pesos da binomial.

Quando  $a$  se refere 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  $a_L$ . Os estimadores REML são dados por:

$$\hat{\sigma}_{a_L}^2 = \frac{\hat{a}_L' A^{-1} \hat{a}_L}{q - tr(A^{-1} C^{22}) / \sigma_{e_L}^2}; \quad \hat{\sigma}_{e_L}^2 = \frac{(y - X\hat{\beta}_L - Z\hat{a}_L)' S^{-1} (y - X\hat{\beta}_L - Z\hat{a}_L)}{N - r(x) - q + tr(A^{-1} C^{22}) / \sigma_{a_L}^2} \sigma_{e_L}^2.$$

O processo iterativo é repetido até a convergência, com o valor predito de  $\hat{\theta} = X\hat{\beta}_L + Z\hat{a}_L$ , transformado, usando a função de ligação para obtenção do novo valor predito de  $\mu$  através de  $\hat{\mu} = \frac{e^{\hat{\theta}}}{1 + e^{\hat{\theta}}}$ , o qual é utilizado para atualização de  $S^{-1}$  e  $y^*$ .

Em resumo, o processo de estimação envolve:

- Estimação de  $\mu = n_1 / N$ , em que  $n_1$  é o número de indivíduos que recebem o escore 1, dentre  $N$  indivíduos avaliados.
- Obtenção de  $y^*$ , a partir de  $y$  e  $\mu$  (neste passo, a variável passa do intervalo (0,1) para a reta real, ou seja, a função é linearizada).
- Estimação de  $\hat{\beta}_L$  e  $\hat{a}_L$ , dados os valores atuais ou correntes de  $\mu$ ,  $\sigma_{e_L}^2$  e  $\sigma_{a_L}^2$ .
- Obtenção de  $\hat{\sigma}_{e_L}^2$  e  $\hat{\sigma}_{a_L}^2$ , iterativamente e, após a convergência, proceder à obtenção atualizada de  $\hat{\beta}_L$  e  $\hat{a}_L$ .
- Obtenção de  $\hat{\eta} = \hat{\theta} = X\hat{\beta}_L + Z\hat{a}_L$ .
- Obtenção de novo valor predito de  $\mu$ , usando a função de ligação, por meio de  $\hat{\mu}_L = \frac{e^{\hat{\theta}}}{1 + e^{\hat{\theta}}}$  (neste passo, a variável volta ao intervalo (0,1)).
- Atualização de  $S^{-1}$  via  $S^{-1} = \hat{\mu}_1(1 - \hat{\mu}_1) \frac{1}{\hat{\sigma}_{e_L}^2}$  e de  $y^*$  via 
$$y^* = \log \left( \frac{\hat{\mu}_1}{1 - \hat{\mu}_1} \right) + \frac{y - \hat{\mu}_1}{\hat{\mu}_1(1 - \hat{\mu}_1)} = \hat{\theta} + \frac{y - \hat{\mu}_1}{\hat{\mu}_1(1 - \hat{\mu}_1)}.$$
- Voltar ao passo (c), enquanto não atingir a convergência.

É interessante notar que este algoritmo é essencialmente hierárquico, havendo, a cada iteração compreendendo os passos de (a) até (h), a necessidade de convergência no passo (d).

Os algoritmos apresentados são do tipo EM (Schall, 1991). Outros algoritmos relatados em literatura para a análise de modelos lineares generalizados mistos (GLMM) são o PQL (*penalized quasi-likelihood*) de Breslow & Clayton (1993) e o IRREML (*iterated re-weighted REML*) de Engel & Keen (1994). Os 3 procedimentos são equivalentes (Keen & Engel, 1997).

Conforme Mc Cullagh & Nelder (1989),  $L(\hat{\mu}, \phi; y)$  equivale ao logaritmo da função verossimilhança maximizada para um valor fixado do parâmetro de dispersão  $\phi$ . Para dados binários,  $\phi$  é conhecido e igual a 1. Entretanto, na prática, os dados podem exibir algum grau de superdispersão ( $\phi > 1$ ) ou subdispersão ( $\phi < 1$ ). Segundo

Lindsey (1997), o fato de  $\phi$  ser diferente de 1 pode ser explicado pelo não atendimento das suposições (eventos independentes, por exemplo) assumidas para que determinado conjunto de dados tenha distribuição binomial. Neste caso, quebra-se a dependência original existente entre a média ( $\mu$ ) e variância  $[\mu(1-\mu)]$  de uma distribuição Bernoulli. Em presença de sub ou superdispersão, uma correção visando à realização de inferências aproximadas refere-se à multiplicação da variância  $\mu(1-\mu)$  por uma estimativa  $\hat{\phi}$  de  $\phi$  (Firth, 1991; Paul & Islam, 1995).

O software AsReml maximiza a função verossimilhança residual para um valor fixado em 1 para a variância residual. Posteriormente essa deve ser tomada como  $\sigma^2 = \pi^2/3 = 3.29$ , que refere-se à variância da distribuição logística. Se usada a opção !DISP ajusta-se o parâmetro de dispersão como variância residual e os demais componentes de variância devem ser multiplicados por essa estimativa, enquanto  $\sigma^2$  deve ser tomado como 3.29.

No Selegen-Reml/Blup a função verossimilhança residual não é maximizada para um valor fixado em 1 para a variância residual. Estima-se diretamente a variância residual que já contempla o parâmetro de dispersão. Essa variância residual e  $\sigma^2 = 3.29$  são usados para ajustar os demais componentes de variância. O algoritmo usado é o da verossimilhança penalizada ou re-weighted Reml iterativo, que equivale também ao EM-Reml de Schall (1991). A taxa de erro no processo iterativo pode ser fixado em  $10^{-4}$  e não  $10^{-5}$ .

Os valores genéticos preditos ( $\hat{\theta} = X\hat{\beta}_L + Z\hat{a}_L$ ) na escala logística podem ser retornados à escala original via a função de ligação, por meio de  $\hat{\mu}_L = \frac{e^{\hat{\theta}}}{1 + e^{\hat{\theta}}}$  (neste passo, a variável volta ao intervalo (0,1)). Ou seja, substituindo-se os valores obtidos de  $\theta$  na função sigmóide  $[\mu = e^{\theta} / (1 + e^{\theta})]$ , produz-se  $\mu$  no intervalo entre 0 e 1.

No modelo 183 os seguintes ajustes podem ser realizados:

a) modelo individual com matriz de parentesco A. Nesse caso os valores genéticos exibidos na primeira seção do arquivo de resultados referem-se a indivíduos e genitores. Nesse caso, deve-se fornecer um arquivo de pedigree contendo as colunas de indivíduo, pai e mãe.

b) modelo com famílias de irmãos completos. Nesse caso os valores genéticos exibidos na primeira seção do arquivo de resultados referem-se a famílias. Os resultados referentes aos indivíduos devem ser lidos na segunda seção do arquivo de resultados, na penúltima coluna sob o nome de a-indIG. Nesse caso, deve-se fornecer um arquivo de pedigree contendo as colunas de famílias, seguida das colunas de pai e mãe zeradas. Nesse caso, a herdabilidade individual no sentido restrito pode ser obtida via  $h^2 = 2 * h_{2a}$ .

c) modelo com famílias de meios irmãos. Nesse caso os valores genéticos exibidos na primeira seção do arquivo de resultados referem-se a famílias. Os resultados

referentes aos indivíduos devem ser lidos na segunda seção do arquivo de resultados, na última coluna sob o nome de a-indMI. Nesse caso, deve-se fornecer um arquivo de pedigree contendo as colunas de famílias, seguida das colunas de pai e mãe zeradas. Nesse caso, a herdabilidade individual no sentido restrito pode ser obtida via  $h^2 = 4 * h_{2a}$ .

d) modelo com clones não aparentados. Nesse caso, os valores genéticos exibidos na primeira seção do arquivo de resultados referem-se a clones. Nesse caso, deve-se fornecer um arquivo de pedigree contendo as colunas de clones, seguida das colunas de pai e mãe zeradas. Nesse caso, a herdabilidade individual no sentido amplo pode ser obtida via  $h^2_g = h_{2a}$ .

e) modelo com tratamentos genéricos usados em estatística experimental. Nesse caso, os resultados exibidos na primeira seção do arquivo de resultados referem-se às médias dos efeitos de tratamentos do fator aleatório do modelo. Os resultados das médias dos efeitos de tratamentos do fator fixo do modelo devem ser obtidos do arquivo .efe. Testes de médias dos efeitos do fator fixo podem ser efetuadas usando a variância residual do modelo ajustado. Os valores na escala logística ( $\hat{\theta} = X\hat{\beta}_L$ : Efeito 2 do arquivo .efe) podem ser transformados à escala original via a função de ligação, por meio de  $\hat{\mu} = \frac{e^{\hat{\theta}}}{1 + e^{\hat{\theta}}}$  (neste passo, a variável volta ao intervalo (0,1)), em que  $e = 2.718$ . No caso do fator aleatório do modelo tem-se  $\hat{\theta} = X\hat{\beta}_L + Z\hat{a} = u + a$  do Selegen. Deve-se fornecer um arquivo de pedigree contendo as colunas de tratamento (fator aleatório), seguida das colunas de pai e mãe zeradas.

Para caracteres com mais de duas categorias (c), pode-se recodificar a variável resposta y, fazendo-a ficar no intervalo entre 0 e 1 e usar o modelo 183. Para isso, faz-se  $y = (c_i / c_{\max})$ .

## **8. Modelo para o BLUP com Genitores, Populações, Progênes e Gerações (SIPPPG) no método Bulk dentro de progênes no Melhoramento de Plantas Autógamas (Modelo 186)**

Considere a situação em que são avaliadas progênes de uma espécie autógama nas gerações F3 a F6. Essas progênes são pertencentes a várias populações. Em cada geração são avaliadas as progênes e também os genitores das populações em experimentos delineados em blocos incompletos.

O modelo a ser ajustado é o seguinte:  $y = Xg + Za + Sp + Wb + Ti + e$ , em que y é o vetor de dados, g é o vetor dos efeitos de geração (assumidos como fixos) somados à média geral, a é o vetor dos efeitos genéticos aditivos de famílias (assumidos como aleatórios), p é o vetor dos efeitos de populações (assumidos como aleatórios), b é o vetor dos efeitos de blocos (assumidos como aleatórios), i é vetor dos efeitos da interação famílias x gerações (aleatórios) e e é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

O ajuste deste modelo é realizado em duas etapas. Inicialmente, são

realizadas análises de gerações individuais pelo modelo 17 a partir das quais são obtidas as médias ajustadas de progênie que serão utilizadas nas análises subsequentes. As médias ajustadas são dadas por  $\bar{y} = \mu + g \frac{(1-r_{gg}^2)}{r_{gg}^2(1-r_{gg}^2)}$ , em que  $g$  é o efeito genotípico de progênie estimado pelo modelo 17,  $r_{gg}^2$  é o quadrado da acurácia seletiva e  $(1-r_{gg}^2)$  pondera pelas diferentes precisões associadas às progênies, sendo  $\frac{(1-r_{gg}^2)}{(1-r_{gg}^2)}$  a média das precisões das avaliações de todas as progênies.

Supondo cinco genitores, no modelo 17 a coluna de efeitos fixos (repetições) deve conter os seguintes códigos: pai1, pai2, pai3, pai4, pai5 e prog, sendo esta última associada a todas as progênies de todas as populações.

Na segunda etapa deve-se usar no modelo 186 as médias ajustadas  $\bar{y}$ . A seguinte sequência de colunas é exigida no arquivo de dados:

*Geração Obs Progênie Geração População Interação Nrep Variáveis*

A primeira coluna, referente às gerações, deve ser codificada com S1, S2, S3 e S4. Se uma ou algumas gerações não estiverem presentes, não há problema. Nesse modelo não é necessária a inclusão dos efeitos dos genitores e das populações em F1 e F2, pois cada um desses efeitos contribuem com menos de 1% para o valor do SIPPFG (Resende et al., 2015). Mas, no modelo 17 são obtidos os efeitos dos genitores para propósitos comparativos com as novas progênies. O ajuste dos genitores como efeitos fixos contribui também para um bom ajuste das médias de progênies no modelo 17.

## 9. Avisos e outras inclusões

### Herdabilidade de médias e Acurácias

As herdabilidades ao nível de médias de repetições e acurácias apresentadas no arquivo .res somente são válidas se o preenchimento das colunas de repetição e de árvore for realizado exatamente conforme o manual. Nos casos em que os modelos forem adaptados, e também de maneira geral, deve-se usar a acurácia do arquivo .fam e obter a herdabilidade ao nível de médias elevando a acurácia ao quadrado.

### Covariável

A correção simples por covariável pode ser feita mais rapidamente e diretamente pelo modelo 119. Use a opção BLUP,  $h^2 = 0$  e variância fenotípica = 1. No .dev já saem os valores corrigidos.

### Modelos de efeitos fixos

Modelos de efeitos fixos, tais quais os fatoriais podem ser ajustados no Selegen usando a opção BLUP e fixando a  $h^2$  e  $c_2$ 's em 1.

**Taxa de autofecundação (S) no Selegen-Reml/Blup (modelos 110 a 112)**

S	F	r sob sistema misto	r sob sistema de alogamia	S no sistema alogamia*
1.0	1.00	2.00	1.00	1.83
0.9	0.82	1.64	0.90	1.56
0.8	0.67	1.35	0.81	1.32
0.7	0.54	1.11	0.72	1.11
0.6	0.43	0.91	0.64	0.91
0.5	0.33	0.75	0.56	0.73
<b>0.4</b>	0.25	0.61	<b>0.50</b>	0.56
<b>0.3</b>	0.18	<b>0.50</b>	0.42	<b>0.41</b>
0.2	0.11	0.40	0.36	0.27
0.1	0.05	0.32	0.30	0.13
0.05	0.03	0.28	0.28	0.05

\*: para uso no Selegen-REML/BLUP

Sob o sistema misto, com S igual a 0.3, as progênies podem ser tratadas como de irmãos completos ( $r = 0.5$  no Quadro acima). O mesmo é válido para o sistema alogamia, com S igual a 0.4 ( $r = 0.5$  no Quadro acima). No Selegen, se o sistema é misto e S é igual a 0.3, deve-se entrar com S igual a 0.41 ( $S = 0.41$  na última coluna da no Quadro acima). Se S é igual a 0.2 deve-se entrar com S igual a 0.27 e assim sucessivamente. Tem-se:

$$r = \frac{1}{2} \frac{(1+S)^2}{2-S} = r_{\text{misto}}; \quad r = \frac{(1+S)^2}{4} = r_{\text{alogamia}}, \text{ conforme Resende (2015). Substituindo}$$

em  $r_{\text{alogamia}}$ , valores de S da última coluna, obtém-se  $r_{\text{misto}}$  da terceira coluna. Essa é a filosofia do software Selegen-Reml/Blup.

**Análise de Repetibilidade**

Mais dois métodos para a estimação do coeficiente de repetibilidade foram incluídos. O método de Rutledge (1974) baseado na análise de componentes principais (PCA) foi programado no final modelo 103 e o método de Mansour et al. (1986) denominado análise estrutural foi incluído no final do modelo 105 como média das correlações fenotípicas. Em ambos os casos o arquivo de dados deve conter as medidas repetidas alocadas em várias colunas de variáveis.

**Interação entre Organismos (Modelo 185)**

Existem também as estimações baseadas em *genótipos de dois organismos em um único indivíduo*, tais quais como ocorre com a seringueira, a laranjeira, o pessegueiro, a mangueira, que são cultivadas via *porta-enxerto + enxerto*. Ocorre também nos experimentos envolvendo interação *planta x patógeno*. Este tipo de análise estatística é contemplada no modelo 185 do software Selegen-Reml/Blup. O estudo da interação planta x patógeno envolve a avaliação de diferentes genótipos (acessos) da espécie vegetal submetidas à inoculações com diferentes raças, estirpes ou fontes de inóculos da espécie de microorganismos (bactérias, fungos, nematoides). Cada planta deve ser inoculada com apenas um tipo da espécie de microorganismo. O experimento deve conter repetições de cada combinação planta-inóculo. Assim, o modelo de análise conterá os efeitos de repetições (b, se o delineamento for em blocos), efeito

genotípico (g) da espécie vegetal, efeito genotípico (m) da espécie de microorganismo, interação genotípica planta x patógeno (gm) e resíduo (e). Esse modelo, para uma variável em um vetor y, é então dado por  $y = Xb + Zg + Wm + Tgm + e$ , em que X, Z, W e T são matrizes de incidência para os respectivos vetores de efeitos. A sequência de colunas no arquivo de dados deve ser: indivíduo, genótipo da planta, blocos, genótipo do patógeno, interação genotípica planta x patógeno, observação, variáveis.

Assim, a causa genética da doença pode ser atribuída a três fatores:

- i. Efeito no fenótipo explicado pelo genótipo do patógeno;
- ii. Efeito no fenótipo explicado pelo genótipo da planta;
- iii. Efeito no fenótipo explicado em conjunto pelo patógeno e pela planta, ou seja, pela combinação genótipo do patógeno – genótipo da planta (é a verdadeira interação planta – patógeno ou capacidade específica de combinação - CEC). Os valores genéticos preditos pelo modelo, para cada combinação genótipo do patógeno – genótipo da planta, são dados por:  $VGC = \hat{u} + \hat{g}_i + \hat{m}_j + \hat{gm}_{ij}$ .

## 10. Novos modelos para Plantas Autógamas (Modelos 187 a 200)

### BLUP com Genitores e Populações F<sub>1</sub> ou F<sub>2</sub> ou F<sub>3</sub> em Plantas Autógamas

Geralmente estes experimentos (baseados em populações na geração F<sub>1</sub> ou F<sub>2</sub> ou F<sub>3</sub>) estão associados a cruzamentos dialélicos entre os genitores e, portanto, permitem estimar as capacidades geral (cgc) e específica (cec) de combinação bem como os respectivos componentes de variância e herdabilidades. Estas análises podem ser obtidas pelos Modelos 189 e 190 do Selegen REML/BLUP.

Para a seleção de indivíduos dentro dessas populações deve ser utilizado o Modelo 24, o qual consiste da soma dos efeitos de população mais os efeitos de indivíduo dentro de população. Esse modelo necessita de uma estimativa da variabilidade genética média dentro de população (Vad) e da correspondente herdabilidade (h<sup>2</sup><sub>ad</sub>). A obtenção desta é realizada via análise (pelo Modelo 187) simultânea dos dados dos genitores e das populações F<sub>2</sub> ou F<sub>3</sub>, conforme escrito a seguir, com base em Resende (2015) e Resende et al. (2016):

- (i) Obtenção da variância genética aditiva original da população de genitores:

$$\sigma_{a0}^2 = \text{Var}(\text{Pais}) / 2;$$

- (ii) Obtenção de F<sub>ST</sub> (F<sub>ST</sub> é o coeficiente de endogamia devido à diferenciação entre

populações): sendo  $\text{Var}(F_2) = 2F_{ST}\sigma_{a0}^2$ , tem-se  $F_{ST} = \frac{\text{Var}(F_2)}{\text{Var}(\text{Pais})}$  ;

- (iii) Obtenção da variância genética média dentro de população (Vad):

$$Vad = (1 - F_{ST})\sigma_{a0}^2;$$

- (iv) Obtenção da herdabilidade média dentro de população (h<sup>2</sup><sub>ad</sub>):

$h_{ad}^2 = (1 - F_{ST})\sigma_{a0}^2 / \text{Var}(\text{Residual}) = Vad / \text{Var}(\text{Residual})$ . Esse valor de h<sup>2</sup><sub>ad</sub> deve ser fornecido ao Selegen no Modelo 24, onde é usado para ponderar o desvio (dev) do valor fenotípico individual corrigido para os efeitos da média geral, de populações e de parcelas. Assim, o índice final calculado pelo Selegen é dado por:  $I = b_{\text{pop}} F_{2\text{pop}} + b_{\text{ind}} \text{dev}$ , sendo F<sub>2pop</sub> o BLUP do efeito de população. Podem ser obtidas também a variância e



herdabilidade total entre e dentro de populações, respectivamente, via:

$$\sigma_{at}^2 = 2F_{ST}\sigma_{a0}^2 + (1 - F_{ST})\sigma_{a0}^2 = (1 + F_{ST})\sigma_{a0}^2 = \text{Var}(F_2) + \text{Vad}; \quad h_{aT}^2 = \frac{(1 + F_{ST})\sigma_{a0}^2}{\sigma_y^2} = \frac{\sigma_{aT}^2}{V_f}$$

O modelo 191 do Selegen permite estimar a variância genética de dominância por meio de  $\sigma_{dom}^2 = \text{Var}(F_1) - \text{Var}(Pais)/2$  e o modelo 188 permite a seleção pelo BLUP de população + progênie dentro de uma dada geração de endogamia. O Selegen para espécies autógamas tem como principais modelos: 24, 59, 60, 61, 98, 180, 181, 186, 187 até 200.

### Índice de Seleção BLUP com Genitores e Populações F<sub>1</sub>, F<sub>2</sub> ou F<sub>3</sub>

Na seleção de populações F<sub>2</sub> geralmente tem sido usada apenas as informações de populações, ignorando-se as informações de genitores. A seguir é descrito um novo método que pode ser aplicado na situação em que são avaliados simultaneamente populações F<sub>2</sub>'s e seus genitores. O índice é da forma  $I_a = \hat{b}_1[(g_1 + g_2)/2] + \hat{b}_2g_{F_2}$ , em que  $(g_1 + g_2)/2$  refere-se à média dos valores genéticos aditivos preditos (com base nos dados dos genitores no experimento) dos genitores e  $g_{F_2}$  equivale ao valor genético aditivo predito (com base nos dados das populações no mesmo experimento) da população F<sub>2</sub>. Os pesos são dados por  $\hat{b}_1 = \frac{(1 - r_{aaF_2}^2)}{1 - r_{aaPais}^2 r_{aaF_2}^2}$  e

$$\hat{b}_2 = \frac{(1 - r_{aaPais}^2)}{1 - r_{aaPais}^2 r_{aaF_2}^2}. \text{ A acurácia do índice é dada por } r_{aal}^2 = \sqrt{1 - \frac{(1 - r_{aaPais}^2)(1 - r_{aaF_2}^2)}{1 - r_{aaPais}^2 r_{aaF_2}^2}}. \text{ Tem-se as}$$

seguintes definições de acurácia visando a geração F<sub>∞</sub>:  $r_{aaPais}^2 = (1/2)(r_{aaMãe}^2 + r_{aaPai}^2)$ : quadrado da acurácia do valor genético predito da média dos dois pais;  $r_{aaMãe}^2$ : quadrado da acurácia do valor genético predito do genitor usado como mãe;  $r_{aaPai}^2$ : quadrado da acurácia do valor genético predito do genitor usado como pai;  $r_{aaF_2}^2$ : quadrado da acurácia do valor genético predito da população F<sub>2</sub>. Essas acurácias são dadas por:  $r_{aaPais}^2 = 0.5\sigma_a^2 / V_{\text{fenot} / \text{pais}}$ ;  $r_{aaF_2}^2 = 2F_{ST}^2\sigma_a^2 / V_{\text{fenot} / F_2}$ ;  $r_{aaF_1}^2 = 0.5\sigma_a^2 / V_{\text{fenot} / F_1}$ . Se  $F_{ST} = 0.5$ ,  $r_{aaPais}^2 = r_{aaF_2}^2 = r_{aaF_1}^2$ . Para populações F<sub>1</sub>, os pesos e acurácia são dados pelas mesmas expressões, trocando-se F<sub>2</sub> por F<sub>1</sub>. Como exemplo considere a avaliação de três caracteres em uma população 1-2. As acurácias  $r_{aaPais}^2$ ,  $r_{aaF_2}^2$  e  $r_{aaF_1}^2$  podem ser calculadas via acurácias do arquivo .fam do Selegen, dividindo-as por 4, 1/F<sub>ST</sub> e 2, respectivamente.

	$r_{aaMãe}^2 / r_{aaMãe}^2$	$r_{aaPai}^2 / r_{aaPai}^2$	$r_{aaPais}^2 / r_{aaPais}^2$	$r_{aaF_2}^2 / r_{aaF_2}^2$ *	F <sub>ST</sub>	$r_{aal}^2 / r_{aal}^2$
Caráter 1	0.34/ 0.12	0.35/ 0.13	0.35/ 0.13	0.16/0.03	0.36	0.39/ 0.15
Caráter 2	0.44/ 0.19	0.44/ 0.20	0.44/ 0.20	0.62/0.38	0.73	0.68/ 0.46
Caráter 3	0.44/ 0.20	0.45/ 0.20	0.45/ 0.20	0.61/0.37	0.68	0.67/ 0.45

\* Acurácia e confiabilidade projetada de populações F<sub>2</sub> como indicadora do valor genético de uma linhagem na geração F infinito.

Verifica-se que as informações dos genitores são tão ou mais importantes que as informações das médias das populações F<sub>2</sub>, dependendo de F<sub>ST</sub>. O F<sub>ST</sub> mede a proporção da variabilidade total (2V<sub>ao</sub>) que é distribuída entre populações. Quanto maior 2F<sub>ST</sub> V<sub>ao</sub>, maior é a eficiência da F<sub>2</sub>. Por sua vez, o índice de seleção combinando as duas fontes de informação é o procedimento ótimo (BLUP) e propicia a maior acurácia ( $r_{aal}$ ) possível. Para o caráter 1, verifica-se que a acurácia do índice (0.39) foi superior a acurácia da F<sub>2</sub> (0.16) e pouco maior que a dos genitores (0.35). Esses resultados são válidos sob modelo com herança completamente aditiva. Outro método para escolha de populações é dado pelo índice



$I_b = F_{2pop_i} + 3.09\sqrt{(1-F_{ST})2\sigma_{a0}^2 Var(Residual_i)/Var(Residual)}$ , em que  $F_{2pop_i}$  é o BLUP do efeito de população  $i$  e o termo dentro da raiz é uma medida da magnitude da variabilidade genética dentro da população  $i$ . A quantidade  $Var(Residual_i)$  refere-se à variabilidade fenotípica dentro da população  $i$  e encontra-se no arquivo .het do Selegen. O valor do índice  $I$  refere-se a uma estimativa do valor genético da melhor linhagem a ser selecionada ao final do processo seletivo. O valor 3.09 equivale ao ponto de truncamento (número de desvios padrões) na curva Normal associado à seleção de 1 em 1000 linhagens. A quantidade 3 desvios padrões abrange 99.7% dos dados de uma distribuição Normal. Como exemplo, considere os seguintes resultados para uma população 2-5:  $F_{2pop_i} = 58.93$ ;  $Var(Residual_i) = 1895.49$ ;  $2\sigma_{a0}^2 = 13.06$ ;  $F_{ST} = 0.3577$ ;  $Var(Residual) = 1173.46$ ; média geral = 56.34. O valor estimado de  $I$  equivale a 75.02 e refere-se ao valor genético da melhor linhagem a ser extraída. Comparando-se com a média geral, tem-se  $75.02/56.34 = 1.33$  e, portanto, o ganho genético será de 33%. Comparando-se com a média da população 2-5, tem-se  $75.02/58.93 = 1.27$  e, portanto, o ganho genético com a seleção dentro da população 2-5 será de 27%. Os melhores índices  $I$  contemplam simultaneamente média alta e ampla variabilidade genética. Um índice melhorado e com o mesmo propósito de  $I_b$ , usa a seleção combinada de genitores e populações  $F_2$  e é dado por:  $I_c = \hat{b}_1[(g_1 + g_2)/2] + \hat{b}_2g_{F_2} + 3.09\sqrt{(1-F_{ST})2\sigma_{a0}^2 Var(Residual)/Var(Residual)}$ . Tal índice contempla genitores, populações  $F_2$  e a variabilidade entre indivíduos dentro de populações  $F_2$ .

### Dialélicos e Fatoriais com Genitores Endógamos e mesmo $F$ diferente de 1

Quando os dois genitores são endógamos com mesma endogamia  $F$  diferente de 1, os componentes de variâncias são:

#### Fatoriais

$\sigma_m^2 = (1/4)(1+F)\sigma_a^2$ : variância entre machos endógamos;

$\sigma_f^2 = (1/4)(1+F)\sigma_a^2$ : variância entre fêmeas endógamas;

$\sigma_{cec}^2 = (1/4)(1+F)^2\sigma_d^2$ : variância da capacidade específica de combinação.

$\sigma_g^2 = \sigma_m^2 + \sigma_f^2 + \sigma_{cec}^2 = [(1/4)(1+F)]\sigma_a^2 + [(1/4)(1+F)]\sigma_a^2 + [(1/4)(1+F)^2]\sigma_d^2$ : variância genotípica total entre famílias de irmãos completos.

$\sigma_a^2 = \{[4/(1+F)]\sigma_m^2 + [4/(1+F)]\sigma_f^2\}/2$ : variância genética aditiva;

$\sigma_d^2 = [4/(1+F)^2]\sigma_{cec}^2$ : variância genética de dominância.

Com famílias  $F_3$  ou  $S_1$ , tem-se  $F = (1/2)$  e:

$$\sigma_a^2 = \{[4/(1+F)]\sigma_m^2 + [4/(1+F)]\sigma_f^2\}/2 = [(8/3)\sigma_m^2 + (8/3)\sigma_f^2]/2$$

$$\sigma_d^2 = [4/(1+F)^2]\sigma_{cec}^2 = (16/9)\sigma_{cec}^2$$

#### Dialélicos

$\sigma_{gen}^2 = \{[(1/4)(1+F) + (1/4)(1+F)]/2\}\sigma_a^2$ : variância entre genitores;

$\sigma_{cec}^2 = (1/4)(1+F)^2\sigma_d^2$ : variância da capacidade específica de combinação;

$\sigma_g^2 = \sigma_{gen}^2 + \sigma_{cec}^2 = \{[(1/4)(1+F) + (1/4)(1+F)]/2\}\sigma_a^2 + [(1/4)(1+F)^2]\sigma_d^2$ : variância genotípica total entre famílias de irmãos completos.

$\sigma_a^2 = \{[4/(1+F)] + [4/(1+F)]/2\} \sigma_{gen}^2 = 4/(1+F) \sigma_{gen}^2$  : variância genética aditiva;  
 $\sigma_d^2 = [4/(1+F)^2] \sigma_{cec}^2$  : variância genética de dominância.

Com famílias F<sub>3</sub> ou S<sub>I</sub>, tem-se  $F = (1/2)$  e:

$$\sigma_a^2 = 4/(1+F) \sigma_{gen}^2 = (8/3) \sigma_{gen}^2$$

$$\sigma_d^2 = [4/(1+F)^2] \sigma_{cec}^2 = (16/9) \sigma_{cec}^2$$

### Dialélicos e Fatoriais com um Genitor Endógamo e outro não Endógamo

Para espécies alógamas (todos os genitores com  $F = 0$ ) e autógamas (todos os genitores com  $F = 1$ ), as covariâncias entre parentes sob cruzamentos dialélicos e fatoriais foram apresentadas no Capítulo 7. Quando um genitor é endógamo (com endogamia  $F$ ) e outro é não endógamo, os componentes de variâncias são:

#### Fatoriais assumindo fêmeas com $F = 1$

$\sigma_m^2 = (1/4) \sigma_a^2$  : variância entre machos não endógamos;  
 $\sigma_f^2 = (1/4)(1+F) \sigma_a^2$  : variância entre fêmeas endógamas;  
 $\sigma_{cec}^2 = (1/4)(1+F) \sigma_d^2$  : variância da capacidade específica de combinação.  
 $\sigma_g^2 = \sigma_m^2 + \sigma_f^2 + \sigma_{cec}^2 = (1/4) \sigma_a^2 + [(1/4)(1+F)] \sigma_a^2 + [(1/4)(1+F)] \sigma_d^2$  : variância genotípica total entre famílias de irmãos completos.  
 $\sigma_a^2 = \{4\sigma_m^2 + [4/(1+F)] \sigma_f^2\} / 2$  : variância genética aditiva;  
 $\sigma_d^2 = [4/(1+F)] \sigma_{cec}^2$  : variância genética de dominância.

Com  $F = 1$ , tem-se:

$$\sigma_f^2 = (1/2) \sigma_a^2;$$

$$\sigma_{cec}^2 = (1/2) \sigma_d^2;$$

$$\sigma_g^2 = \sigma_m^2 + \sigma_f^2 + \sigma_{cec}^2 = (3/4) \sigma_a^2 + (1/2) \sigma_d^2;$$

$$\sigma_a^2 = (4\sigma_m^2 + 2\sigma_f^2) / 2 = 2\sigma_m^2 + \sigma_f^2;$$

$$\sigma_d^2 = 2\sigma_{cec}^2.$$

#### Dialélicos

$\sigma_{gen}^2 = \{[(1/4) + (1/4)(1+F)]/2\} \sigma_a^2$  : variância entre genitores;  
 $\sigma_{cec}^2 = (1/4)(1+F) \sigma_d^2$  : variância da capacidade específica de combinação;  
 $\sigma_g^2 = \sigma_{gen}^2 + \sigma_{cec}^2 = \{[(1/4) + (1/4)(1+F)]/2\} \sigma_a^2 + [(1/4)(1+F)] \sigma_d^2$  : variância genotípica total entre famílias de irmãos completos.  
 $\sigma_a^2 = \{[4 + [4/(1+F)]]/2\} \sigma_{gen}^2$  : variância genética aditiva;  
 $\sigma_d^2 = [4/(1+F)] \sigma_{cec}^2$  : variância genética de dominância.

Com  $F = 1$ , tem-se:

$$\sigma_{gen}^2 = (3/4) \sigma_a^2;$$

$$\sigma_{cec}^2 = (1/2) \sigma_d^2;$$

$$\sigma_g^2 = \sigma_{gen}^2 + \sigma_{cec}^2 = (3/4) \sigma_a^2 + (1/2) \sigma_d^2.$$

$$\sigma_a^2 = (4/3)\sigma_{gen}^2 ;$$

$$\sigma_d^2 = 2\sigma_{cec}^2 .$$

O Selegen implementa os seguintes modelos após o modelo 186:

<b>Modelo</b>	<b>Objetivo</b>	<b>Modelo</b>	<b>Objetivo</b>
187	Avaliação de Populações F <sub>2</sub> (ou F <sub>3</sub> ) e Genitores	194	Seleção Simultânea pelos Genitores, F <sub>1</sub> e F <sub>2</sub> (Cruz. Fatoriais)
188	Avaliação de Populações e Progenies/Pop em uma Geração e Locais	195	Seleção com 4 entidades genéticas: spp, pop, prog, ind (com med. repet.)
189	Avaliação de Dialélico de Linhagem	196	Fatorial, Blocos, Interpopulacional para Genitores com F entre o e 1
190	Avaliação de Dialélico de Linhagem em vários locais	197	Fatorial, em Látice, Interpopulacional para Genitores com F entre o e 1 e Interação G x E
191	Avaliação de Populações F <sub>1</sub> e Genitores	198	Dialélico entre Genitores com F entre o e 1
192	Cruzamentos Fatoriais entre Linhagens: Dados Simultâneos de Populações e Genitores	199	Dialélico entre Genitores com F entre o e 1 com Interação G x E
193	Progenies de várias populações, com medidas repetidas (C. arábica)	200	Dialelico em Alógamas com Medidas Repetidas

Essas modelos encontram-se nas janelas “Espécie Perenes Autógamas”, “Dialélicos Genitores não Aparentados”, “Fatoriais Interpopulacionais”.

Os arquivos de dados devem conter:

*Modelo 187: Ind Pop Fixo Parc Pais Planta Variáveis*

*Modelo 188: Ind Pop Fixo Interação Prog Planta Variáveis*

*Modelo 189: Ind Pai Mãe Rep Parc Prog Planta Variáveis*

*Modelo 190: Local Ind Pai Mãe Rep Interação Prog Parc Planta Variáveis*

*Modelo 191: Ind PopF<sub>1</sub> Fixo Parc Pais Planta Variáveis*

*Modelo 192: Ind Macho Fixo Fêmea Prog Pais Bloc Planta Variáveis*

*Modelo 193: Ind Prog Fixo Parc Pop Bloc Interação Planta Variáveis*

*Modelo 194: Ind Pai Fixo Pai<sub>2</sub> PopF<sub>1</sub> PopF<sub>2</sub> Bloc Planta Variáveis*

*Modelo 195: Obs Prog Fixo Pop Ind Espécie Bloc Planta Variáveis*

*Modelo 196: Ind Pai Rep Mãe Prog Planta Variáveis*

*Modelo 197: Ind Machos Repetição Fêmeas Prog Interação Bloco Árvore Variáveis*

*Modelo 198: Ind Pai Mãe Rep Parc Prog Planta Variáveis*

*Modelo 199: Local Ind Pai Mãe Rep Interação Prog Parc Planta Variáveis*

*Modelo 200: Medição Ind Pai Mãe Rep Interação Prog Parc Planta Variáveis*

Os modelos 187, 191, 192 e 194 ajustam simultaneamente um modelo para cruzamentos e outro para genitores (que geralmente são incluídos nos experimentos como testemunhas). Nesses casos, torna-se necessário uma codificação especial na coluna “Fixo”, separando cruzamentos e genitores. Para o modelo 192 tem-se:

**Para dados dos cruzamentos:**

$y = Xc + Zm + Wf + Sd + Tb + e$ , em que  $c$  é o efeito fixo da média geral dos cruzamentos;  $m$  é o vetor dos efeitos das capacidades gerais de combinação dos genitores masculinos (aleatórios);  $f$  é o vetor dos efeitos das capacidades gerais de combinação dos genitores femininos (aleatórios);  $d$  é o vetor dos efeitos das capacidades específicas de combinação dos cruzamentos (aleatórios);  $b$  é o vetor dos efeitos de blocos (aleatórios);  $e$  é o vetor dos efeitos do erro (aleatórios).

**Para dados dos genitores:**

$y = Xu + Hg + Tb + e$ , em que  $u$  é o efeito fixo da média geral dos genitores;  $g$  é o vetor dos efeitos genéticos aditivos dos genitores (aleatórios);  $b$  é o vetor dos efeitos de blocos (aleatórios);  $e$  é o vetor dos efeitos do erro (aleatórios).

No modelo 195, os pesos  $b_1$ ,  $b_2$ ,  $b_3$  e  $b_4$  ponderam os efeitos  $g_i$  de progênes, populações, indivíduos e espécies, respectivamente. E, no caso,  $Vg_1$ ,  $Vg_2$ ,  $Vg_3$  e  $Vg_4$  são as variâncias genotípicas entre progênes, populações, indivíduos e espécies, respectivamente. A seleção (apresentada ao final do arquivo .res) é baseada no valor genotípico total dado por  $b_1 \times g_1 + b_2 \times g_2 + b_3 \times g_3 + b_4 \times g_4$ . Os pesos  $b_1$ ,  $b_2$ ,  $b_3$  e  $b_4$  são fornecidos via tela e podem ser tomados como  $b_1 = b_2 = b_4 = 1$  e  $b_3 = f(Vg_1)/Vg_3$ , em que  $f(Vg_1)$  é uma função da variação genética entre os tipos (irmãos germanos:  $f(Vg_1) = Vg_1$ ; meios irmãos:  $f(Vg_1) = 3/4 Vg_1$ ; progênes).

O modelo 194 é similar e refere-se ao índice de seleção com pais,  $F_1$ 's e  $F_2$ 's em plantas autógamas. O índice é da forma  $I = \hat{b}_g[(g_1 + g_2)/2] + \hat{b}_{F_2}g_{F_2} + \hat{b}_{F_1}g_{F_1}$ . Os pesos são dados por  $\hat{b}_g = \frac{(1-r_{aaF_1}^2)(1-r_{aaF_2}^2)}{1-r_{aaPais}^2-r_{aaF_1}^2-r_{aaF_2}^2}$ ;  $\hat{b}_{F_2} = \frac{(1-r_{aaPais}^2)(1-r_{aaF_1}^2)}{1-r_{aaPais}^2-r_{aaF_1}^2-r_{aaF_2}^2}$  e  $\hat{b}_{F_1} = \frac{(1-r_{aaPais}^2)(1-r_{aaF_2}^2)}{1-r_{aaPais}^2-r_{aaF_1}^2-r_{aaF_2}^2}$ . A acurácia do índice é dada por  $r_{aal} = \sqrt{1 - \frac{(1-r_{aaPais}^2)(1-r_{aaF_1}^2)(1-r_{aaF_2}^2)}{1-r_{aaPais}^2-r_{aaF_1}^2-r_{aaF_2}^2}}$ . Resultados são apresentados a seguir, supondo  $r_{aaF_1} = r_{aaPais}$ . No caso,  $bg = b_1$ ;  $b_{F_1} = b_2$ ;  $b_{F_2} = b_3$ .

	$r_{aaPais}$	$r_{aaF_2}$	$F_{ST}$	$r_{aaIPais \times F_2}$	$r_{aaIPais \times F_2 - F_1}$
Caráter 1	0.35	0.16	0.36	0.39	0.52
Caráter 2	0.44	0.62	0.73	0.68	0.77
Caráter 3	0.45	0.61	0.68	0.67	0.77

## 11. Referências

RESENDE, M. D. V. de; HIGA, A. R.; LAVORANTI, O. J. Predição de valores genéticos no melhoramento de *Eucalyptus* – melhor predição linear (BLP). In: Congresso Florestal Brasileiro, 7, 1993, Curitiba. **Anais...**, Curitiba: SBS, 1993. p. 144-147.

RESENDE, M. D. V. de; 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*. *Pesquisa Florestal Brasileira*, Colombo, n.32/33, p.18-45, 1996.

RESENDE, M. D. V. de; ROSA-PEREZ, J. R. H. **Genética Quantitativa e Estatística no Melhoramento Animal**. Curitiba: Imprensa Universitária - UFPR, 1999. 496 p.

RESENDE, M. D. V. **Inferência Bayesiana e simulação estocástica (Amostragem de Gibbs) na estimação de componentes de variância e valores genéticos em plantas perenes**. Colombo: Embrapa Florestas, 2000. 68 p.

RESENDE, M.D.V. **Genética Biométrica e Estatística no Melhoramento de Plantas Perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975p.

RESENDE, M. D. V. **Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético**. Colombo: Embrapa Florestas, 2007. 561 p.

RESENDE, M. D. V. de. **Selegen-Reml/Blup: Sistema Estatístico e Seleção Genética Computadorizada via Modelos Lineares Mistos**. Colombo: Embrapa Florestas, 2007. 360 p.

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

RESENDE, M. D. V. **Genômica Quantitativa e Seleção no Melhoramento de Plantas Perenes e Animais**. Colombo: Embrapa Florestas, 2008. 330 p.

RESENDE, M.D.V.; SILVA, F.F.; LOPES, P.S.; AZEVEDO, C.F. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa/Departamento de Estatística. 2012. 291 p. [http://www.det.ufv.br/ppestbio/corpo\\_docente.php](http://www.det.ufv.br/ppestbio/corpo_docente.php)

RESENDE, M. D. V. de; RAMALHO, M.A.P.; GUILHERME, S. R.; ABREU, A.F.B. Multigeneration index in the within progenies bulk method for breeding of self-pollinated plants. *Crop Science*, 2014.

RESENDE, M. D. V.; RAMALHO, M. A. P.; CARNEIRO, P.C.S.; CARNEIRO, J.E.S.; BATISTA, L. G.; GOIS, I. B. Selection index with parents, populations, progenies and generations effects in autogamous plant breeding. *Crop Science*, 2015.

RESENDE, M.D.V.; SILVA, F.F. **Estatística Matemática, Biométrica e Computacional: Modelos Mistos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência**. Viçosa: Universidade Federal de Viçosa/Departamento de Estatística. 2014. 888 p.

RESENDE, M.D.V. **Genética Quantitativa e de Populações**. Viçosa: Universidade Federal de Viçosa/Departamento de Estatística, 2015. 422p.

@@