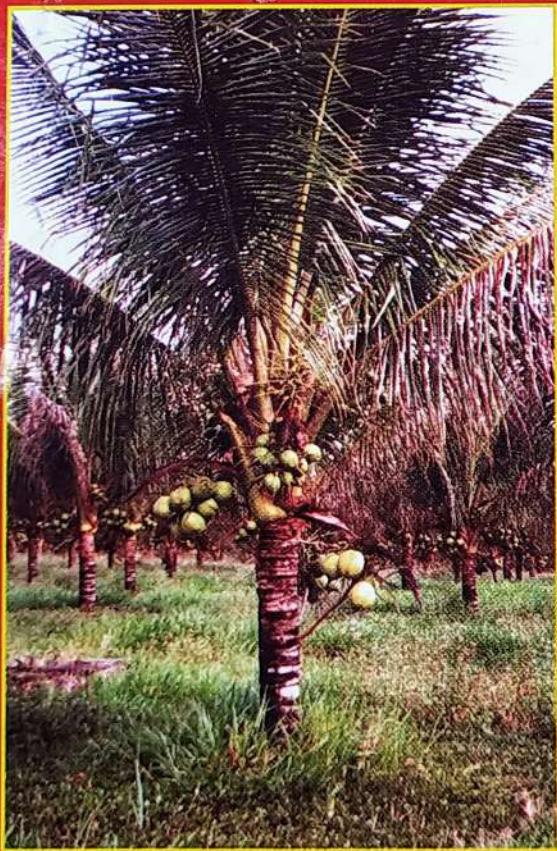


GENÉTICA BIOMÉTRICA e ESTATÍSTICA

no

Melhoramento de Plantas Perenes



Marcos Deon Vilela de Resende

Embrapa

Genética Biométrica

e

Estatística

no

Melhoramento de Plantas Perenes

República Federativa do Brasil

Fernando Henrique Cardoso
Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Marcus Vinicius Pratini de Moraes
Ministro

Empresa Brasileira de Pesquisa Agropecuária – Embrapa

Conselho de Administração

Márcio Fortes de Almeida
Presidente

Alberto Duque Portugal
Vice-Presidente

Dietrich Gerhard Quast
José Honório Accarini
Sérgio Fausto
Urbano Campos Ribeiral
Membros

Diretoria-Executiva da Embrapa

Alberto Duque Portugal
Diretor-Presidente

Bonifacio Hideyuki Nakasu
Dante Daniel Giacomelli Scolari
José Roberto Rodrigues Peres
Diretores-Executivos

Embrapa Florestas

Vitor Afonso Hoeflich
Chefe-Geral

Moacir José Sales Medrado
Chefe-Adjunto de Pesquisa e Desenvolvimento

João Antônio Pereira Fowler
Chefe-Adjunto de Administração

Erich Gomes Schaitza
Chefe-Adjunto de Comunicação, Negócios e Apoio

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Florestas
Ministério da Agricultura, Pecuária e Abastecimento*

Genética Biométrica

e

Estatística

no

Melhoramento de Plantas Perenes

Marcos Deon Vilela de Resende

*Embrapa Informação Tecnológica
Brasília, DF
2002*

Exemplares desta publicação podem ser solicitados na:

Embrapa Informação Tecnológica

*Parque Estação Biológica – PqEB – Av. W3 Norte (final)
CEP 70770-901 – Brasília, DF
Caixa Postal 040315
Fone: (61) 448-4236
Fax: (61) 340-2753
vendas@spi.embrapa.br
www.sct.embrapa.br*

Embrapa Florestas

*Estrada da Ribeira, Km 111
CEP 83411-000 – Colombo, PR
Caixa Postal 319
Fone: (41) 666-1313
Fax: (41) 666-1276
sac@cnpf.embrapa.br*

Embrapa Informação Tecnológica

*Coordenação editorial
Walmir Luiz Rodrigues Gomes
Mayara Rosa Carneiro*

*Revisão de texto
Vitória Rodrigues e Norma Azeredo*

*Projeto gráfico e editoração eletrônica
Júlio César da Silva Delfino*

*Capa
Carlos Eduardo Felice Barbeiro*

*Tratamento das ilustrações
Mário César Moura de Aguiar*

Embrapa Florestas

*Normalização bibliográfica
Lídia Woronoff*

*Fotos da capa
Veracel Celulose (*Eucalipto*), Paulo de Souza Gonçalves (*Seringueira*),
Neuza Campelo (*Guaraná*), Márcio Henrique Pereira Barbosa (*Cana*),
Cacilda Borges do Valle (*Panicum*), Aymbiré Fonseca (*Café robusta*),
Arquivos da Embrapa (*Demais fotos*)*

1ª edição

1ª impressão (2002): 1000 exemplares

Todos os direitos reservados.

*A reprodução não-autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).*

**CIP - Brasil. Catalogação-na-publicação.
Embrapa Informação Tecnológica.**

Resende, Marcos Deon Vilela de.

*Genética biométrica e estatística no melhoramento de plantas perenes / Marcos
Deon Vilela de Resende. – Brasília : Embrapa Informação Tecnológica, 2002.
975 p.*

*Inclui bibliografia
ISBN 85-7383-xxxxxx*

*1. Planta Perene – Melhoramento genético. 2. Melhoramento genético vegetal.
3. Genética biométrica. 4. Planta – Estatística – Matemática – Método. I. Título.*

CDD 581.15 (21. ed.)

© Embrapa 2002

Dedicatória

*Ao Professor Dr. Roland Vencovsky,
pela escola de Genética Biométrica
que fundou e consolidou no Brasil.*

*A Annelise e a Anna Carolina,
minhas filhas e melhor parte de
mim.*

Apresentação

Em meados de 1999 o autor desta obra terminou seu doutorado em Genética, na Universidade Federal do Paraná. A banca examinadora, em vista da qualidade e profundidade da tese, recomendou que o autor a convertesse num livro, pois, assim, seu conteúdo seria mais difundido entre nós. É, portanto, com grande satisfação que vemos agora surgir este livro. Ele contém mais do que a tese; é mais abrangente, detalhado e aprofundado e encerra o resultado de um grande esforço intelectual, quase todo ele de autodidatismo.

Esta obra será de muita utilidade para os alunos de pós-graduação, os pesquisadores e os docentes de nossas universidades, atuantes na área. Fazendo uma conexão entre a Genética Biométrica, a Estatística, a Experimentação e o Melhoramento Genético, ela tem uma abrangência mais ampla do que pode sugerir seu título e, certamente atenderá, não só os interessados em espécies perenes, mas também os que pesquisam outras categorias de plantas. Sendo suficientemente detalhada, servirá como fonte importante de orientação para a elaboração de programas alternativos ou específicos de informática.

Somente os profissionais com experiência em estatística e álgebra de matrizes terão condições de usufruir seu conteúdo, completamente. Todavia, os vários exemplos de aplicação da teoria apresentados, com seus respectivos processos de análise, são excelente material de estudo e utilização por parte daqueles menos avançados nessas duas áreas. Além disso, contém uma relação extensa de referências bibliográficas que, por si só, será de grande proveito para os estudiosos.

Nos seus vinte e um capítulos, o autor preocupou-se em apresentar conceitos básicos de Genética Quantitativa, Genética de Populações, de Melhoramento Genético e de Biometria, antes de enveredar pelos caminhos da sua aplicação mais aprofundadamente. Para alargar seu escopo, cobre não só as espécies alógamas mas também as autógamas, as com sistema misto de reprodução sexuada e as que permitem a reprodução vegetativa. Esta subdivisão das espécies é, naturalmente, acompanhada dos respectivos procedimentos biométricos em função dos métodos de melhoramento que se aplicam, em cada caso. Como é dirigido à exploração eficiente e completa de dados experimentais, no contexto de programas de seleção, este trabalho não se restringe à análise de caracteres isoladamente, pois enfoca, também, a repetição de observações no tempo, bem como o emprego de vários caracteres, simultaneamente. Além das variáveis contínuas, inclui aquelas que, pela sua expressão, são de natureza descontínua ou discreta. Em função da origem e essência dos materiais genéticos sob avaliação, considera os modelos biométricos aleatórios e fixos.

Quanto ao Melhoramento Genético, que é, do ponto de vista biométrico, o cerne do livro, é feita uma cobertura abrangendo a conservação de germoplasmas, a seleção de populações-base e de cruzamentos, o

melhoramento de híbridos interpopulacionais, a seleção clonal e, naturalmente, o melhoramento intrapopulacional, via seleção recorrente. Neste último item, mostra processos visando comparar métodos seletivos, considera a seleção direta e a indireta, bem como os procedimentos ótimos para maximizar o progresso genético, toca a questão dos limites seletivos e apresenta e utiliza seis aplicativos computacionais (Asreml, Dfreml, Matlab, Mtdfreml, Mtgsam e Selegen) que permitem a seleção computadorizada e a estimativa de componentes de variância. Neste último contexto, é muito interessante a predição do valor genético dos indivíduos, fundamental para a seleção incorporando todas as informações genéticas e não genéticas disponíveis num conjunto de dados, com destaque para os índices multiefeitos e para os melhores preditores lineares não viciados (BLUP).

O conceito de tamanho efetivo populacional e suas aplicações ao melhoramento de espécies alógamas, autógamas e de sistema misto de reprodução são considerados em vários capítulos, com os objetivos de maximizar o ganho genético no curto, médio e longo prazos e de promover e monitorar a conservação genética.

Recebem atenção especial os processos de estimativa de parâmetros genéticos, especialmente componentes de variância, fundamentais para avaliar a eficiência de métodos seletivos, estimar parâmetros e fazer previsões. Por isso, além de tratar o assunto do ponto de vista da Biometria e da Genética, inclui itens sobre delineamentos genéticos e experimentais. Ênfase foi dada aos modelos mistos através do procedimento de máxima verossimilhança restrita (REML) para modelos lineares normais e de regressão aleatória, abordando também a técnica de modelos não lineares mistos via modelos lineares generalizados e a análise espacial de experimentos.

Não se atém apenas aos métodos tradicionais ou freqüentistas de inferência, mas introduz também as técnicas bayesianas, as quais, utilizando dados experimentais, aprimoram as informações previamente existentes sobre os parâmetros e melhoram a confiabilidade das estimativas. Também introduz as técnicas de reamostragem, destinadas principalmente à obtenção de desvios padrões de estimativas e intervalos de confiança associados aos respectivos parâmetros. Mas não foi somente nessa área biométrica que o autor procurou indicar as tendências mais recentes de pesquisa. Da mesma forma, preocupou-se em incluir informações de marcadores genéticos para assistir na seleção de genótipos superiores.

A descrição que fizemos é certamente superficial e incompleta. Apenas uma leitura cuidadosa permitirá verdadeiramente aquilar a contribuição que esta obra dará às áreas a que se destina. Vale, como observação final, lembrar as afirmações, muito ouvidas nos últimos anos, vindas de vários pesquisadores, dizendo que a Genética Quantitativa estava já exaurida. Este livro é garantia e testemunho de que isso não é verdade.

Piracicaba, abril de 2001.
Roland Vencovsky

Sumário

Notação Matemática e Estatística	25
Capítulo Introdutório – Espécies Perenes, Sistemas Reprodutivos e Metodologias de Seleção.....	29
1. Conceitos Básicos de Genética Quantitativa e Parâmetros Genéticos	37
1.1. Base Genética dos Caracteres Quantitativos	38
1.2. Modelo Genético e Fenotípico.....	38
1.3. Componentes da Variação Fenotípica.....	39
1.4. Covariância Genética entre Parentes	40
1.5. Correlação Genética entre Parentes e Correlação Fenotípica Intraclasse.....	40
1.6. Regressão, Correlação e Coeficiente de Determinação Envolvendo os Valores Genéticos e Fenotípicos.....	41
1.7. Parâmetros Genéticos Populacionais.....	42
1.7.1. Conceitos	42
1.7.2. Herdabilidade	43
1.7.3. Repetibilidade individual	44
1.7.4. Correlação entre caracteres	45
1.7.5. Parâmetros genéticos e seleção para características de limiar	46
1.7.5.1. Herdabilidade e seleção	46
1.7.5.2. Correlações fenotípicas e genéticas	50
1.7.5.3. Repetibilidade.....	51
1.8. Métodos Tradicionais de Estimação de Parâmetros Genéticos	52
1.8.1. Análise de regressão.....	52
1.8.2. Análise de variância ou de correlação intraclasse	53
1.8.3. Estimação da herdabilidade a partir de delineamentos clássicos	54
1.8.4. Estimação do coeficiente de variação genética	56
1.9. Intereração Genótipo x Ambiente	56
1.9.1. Conceitos	56
1.9.2. Correlação genética através dos ambientes e número de locais de experimentação	58
1.9.3. Estabilidade dos valores genotípicos preditos	61
1.10. Tamanho Efetivo Populacional e Endogamia	63
1.10.1 Conceito fundamental.....	63
1.10.2. Diferentes conceitos do tamanho efetivo	64
1.10.3. Conceitos aplicados do tamanho efetivo	65
1.10.4. Conceito matemático do tamanho efetivo	66
1.10.5. Tamanho efetivo com sobreposição de gerações.....	67

1.10.6. Endogamia devida ao pequeno tamanho efetivo	67
1.10.7. Efeitos da endogamia na média (depressão endogâmica) e variâncias populacionais.....	69
1.10.8. Expressões do N_e úteis ao melhoramento e conservação genética de espécies perenes	69
1.11. Heterose	71
1.11.1. Conceitos	71
1.11.2. Tipos de heterose.....	73
2. Progresso Genético e Métodos de Seleção.....	75
2.1. Conceitos	76
2.2. Valores Genéticos, Fenotípicos e Seleção	76
2.2.1. Inferências pontuais e intervalares sobre valores genéticos	78
2.3. Equação Geral do Progresso Genético	78
2.3.1. Diferencial de seleção e intensidade de seleção	82
2.3.2. Desvio padrão genético aditivo.....	82
2.3.3. Intervalo entre gerações.....	83
2.3.4. Acurácia, confiabilidade e variância do erro de predição	83
2.4. Métodos de Seleção: Coeficientes de Ponderação das Informações Fenotípicas e Acurárias	86
2.4.1. Seleção individual	86
2.4.2. Seleção de famílias	87
2.4.3. Seleção de irmãos.....	87
2.4.4. Seleção dentro de famílias	88
2.4.5. Seleção com base na progénie	88
2.4.6. Seleção com base nos valores genéticos preditos dos genitores	89
2.4.7. Seleção entre e dentro de famílias	89
2.4.8. Seleção combinada	89
2.4.9. Seleção combinada com base em avaliações repetidas em cada indivíduo da família de meios-irmãos	90
2.4.10. Seleção pela metodologia de modelos mistos: BLUP sob modelo de plantas individuais	90
2.4.11. Predição de valores genotípicos por seleção combinada	91
2.5. Comparação de Métodos de Seleção.....	91
2.6. Correlação Intraclasse entre Valores Genéticos Preditos.....	96
2.7. Tendência Genética e Herdabilidade Realizada	97
2.8. Efeitos da Seleção na Variância, nos Parâmetros Genéticos e no Tamanho Efetivo	98
2.9. Variância da Resposta à Seleção.....	99
3. Seleção Indireta e Uso de Caracteres Auxiliares	103
3.1. Resposta Correlacionada	104
3.2. Co-herdabilidade e Coeficiente de Predição Genética.....	105

3.3. Seleção Indireta	105
3.4. Seleção Utilizando Caracteres Auxiliares ao Melhoramento.....	106
3.5. Herdabilidade Multivariada e Coeficiente de Predição Genética Multivariado.....	106
3.6. Índice de Seleção Biológico e Herdabilidade Generalizada.....	109
3.7. Seleção em Múltiplos Estágios.....	111
4. Sistemas de Seleção para Várias Características.....	113
4.1. Conceitos	114
4.2. Seleção em Tandem	115
4.3. Níveis Independentes de Eliminação.....	116
4.4. Índice de Seleção.....	117
4.5. Comparação entre os Três Sistemas de Seleção.....	119
4.6. Índice de Seleção Restrito.....	120
4.7. Índice de Seleção Combinado.....	122
4.8. Índice de Seleção Combinado com Várias Avaliações por Indivíduo.....	123
4.9. Índice de Seleção Utilizando Valores Genéticos Preditos.....	124
4.10. Uso de Parâmetros Genéticos Padrões para Cômputo de Índices de Seleção	125
4.11. Pesos Econômicos dos Caracteres	126
4.12. Objetivos do Melhoramento e Critérios de Seleção	127
4.13. Exemplo Aplicado Envolvendo a Seleção de Dois Caracteres	128
5. Avaliação Genética pela Metodologia Tradicional.....	133
5.1. Teste de Progênie de Meios-irmãos.....	134
5.1.1. Estimativas de parâmetros genéticos e fenotípicos	134
5.1.2. Métodos de seleção	137
5.1.3.Tamanho efetivo populacional e número efetivo de famílias selecionadas	146
5.1.4.Tamanho efetivo em função da proporção de utilização dos indivíduos	148
5.1.5. Endogamia da população melhorada e correção do ganho genético....	151
5.1.6.Variância do ganho genético e número de indivíduos a serem selecionados	153
5.1.7. Restrição ao tamanho efetivo e melhoramento no longo prazo	155
5.1.8. Variância das estimativas dos parâmetros genéticos	156
5.2. Testes Clonais	158
5.3. Medidas Repetidas.....	160
6. Repetibilidade e Avaliação Genética com Medidas Repetidas....	167
6.1. Conceitos	168

6.2. Repetibilidade e Predição de Valores Genéticos, Genotípicos e Fenotípicos Permanentes.....	168
6.3. Repetibilidade e Predição de Valores Genéticos e Genotípicos em Testes de Progêneres e Clonais.....	171
6.4. Eficiência Seletiva e Número de Medições por Indivíduo	173
6.4.1. Seleção individual	173
6.4.2. Vários métodos de seleção e sistemas de propagação	179
6.5. Repetibilidade de Caracteres com Distribuição Binomial.....	185
6.6. Estimação da Repetibilidade pelo Método de Quadrados Mínimos	186
6.6.1. Amostra de dados não associada à estrutura experimental	186
6.6.2. Dados associados ao delineamento experimental de blocos ao acaso	187
6.7. Estimação da Repetibilidade pelo Método de Máxima Verossimilhança Restrita (REML).....	196
6.8. Estimação Comparativa da Repetibilidade por Diferentes Modelos.....	199
6.9. Estimativas de Repetibilidade em Espécies Perenes e Classificação de Magnitudes.....	206
6.10. Tamanho Amostral e Precisão das Estimativas de Repetibilidade.....	208

7. Procedimentos Ótimos de Seleção com Dados Balanceados e Desbalanceados.....

7.1. Introdução	210
7.2. Predição de Variáveis Aleatórias e Metodologia de Modelos Mistos	210
7.2.1. Predição de variáveis aleatórias	211
7.2.2. BP	211
7.2.3. BLP	212
7.2.4. BLUP	212
7.2.5. Modelos lineares e predição de valores genéticos	213
7.2.5.1. Modelo Linear de Efeitos Fixos	214
7.2.5.2. Modelo Linear de Efeitos Aleatórios	216
7.2.5.3. Modelos Lineares Mistos com Efeitos Fixos e Aleatórios.....	221
7.2.6. Preditores BLP versus BLUP.....	225
7.2.7. Álgebra linear numérica e aspectos computacionais da metodologia de modelos mistos.....	226
7.2.8. Derivações alternativas do BLUP	227
7.2.9. Matriz de parentesco genético aditivo e de dominância.....	228
7.2.9.1. Matriz de parentesco genético aditivo e inversa	228
7.2.9.2. Matriz de parentesco de dominância	231
7.2.10. Acurácia das avaliações genéticas e número efetivo de informações	232
7.3. BLUP Individual e Índice Multiefeitos com Uma Avaliação por Indivíduo	233
7.3.1. Seleção por valores genéticos aditivos (propagação sexuada)	233
7.3.2. Seleção por valores genotípicos (propagação assexuada).....	237

7.3.3. Teste clonal no delineamento de blocos ao acaso com várias plantas por parcela.....	238
7.3.4. Teste de progênie no delineamento de blocos ao acaso com uma planta por parcela: seleção por valores genéticos aditivos	239
7.3.5. Exemplo prático.....	240
7.3.6. Definição do bloco como principal efeito fixo associado às avaliações genéticas em plantas perenes	245
7.4. BLUP Individual e Índice Multiefeitos com Medidas Repetidas	246
7.4.1. Seleção por valores genéticos aditivos (propagação sexuada)	247
7.4.2. Seleção por valores genotípicos (propagação assexuada).....	250
7.4.3. Exemplo prático.....	251
7.5. BLUP Individual e Índice Multiefeitos em Presença de Interação Genótipo x Ambiente	257
7.5.1. Seleção por valores genéticos aditivos (propagação sexuada)	259
7.5.2. Seleção por valores genotípicos (propagação assexuada).....	264
7.5.3. Exemplo prático.....	267
7.5.4. Heterogeneidade de variâncias	271
7.6. BLUP Individual e Índice Multiefeitos em Presença de Interação Genótipo x Ambiente e Medidas Repetidas.....	274
7.6.1. Seleção por valores genéticos aditivos (propagação sexuada)	276
7.6.2. Seleção por Valores genotípicos (propagação assexuada)	281
7.6.3. Exemplo prático.....	284
7.7. BLUP Individual e Índice Multiefeitos Associados a Cruzamentos Fatoriais e Dialélicos	290
7.7.1. Modelo Estatístico	290
7.7.2. Índice multiefeitos para os efeitos aditivos.....	291
7.7.3. Índice multiefeitos para os efeitos de dominância.....	295
7.7.4. BLUP para os efeitos aditivos e de dominância	296
7.7.5. Índice multiefeitos e BLUP para efeitos genotípicos	297
7.7.6. Componentes de variância e dados utilizados.....	297
7.8. BLUP Individual e Índice Multiefeitos Associados ao Látice	300
7.8.1. Delineamento em látice e obtenção de médias ajustadas	301
7.8.2. BLUP e índice multiefeitos ajustando bloco dentro de repetição como efeito fixo.....	304
7.8.3. BLUP ajustando bloco dentro de repetições como efeito aleatório	306
7.8.4. Acurácia e variância do erro de predição dos valores genéticos	308
7.8.5. Análise considerando o delineamento de blocos ao acaso.....	309
7.8.6. Exemplo prático.....	309
7.8.7. Modelos alternativos do BLUP associado ao látice	312
7.9. BLUP Individual e Índice Multiefeitos Envolvendo Diferentes Populações	318
7.9.1. BLUP individual para a predição de efeitos aditivos	322
7.9.2. Teste de procedências ou populações	323
7.9.3. Exemplo prático.....	326
7.10. BLUP Individual e Índice Multiefeitos na Seleção Simultânea de Caracteres	330

7.11. BLUP e Índice de Seleção na Seleção Auxiliada por Marcadores	337
7.12. Seleção Baseada em Médias Harmônicas	344
7.13. Avaliação Genética em Espécies Autopoliplóides	345

8. Variáveis Binomiais e Categóricas: Componentes de Variância e Valores Genéticos..... 349

8.1. Introdução	350
8.2. Estimação da Herdabilidade	350
8.2.1. Escala binomial	350
8.2.2. Escala normal	351
8.2.3. Análise via modelos não lineares e princípios bayesianos	351
8.3. Estimação da Repetibilidade	353
8.3.1. Escala binomial	353
8.3.2. Escala normal	354
8.3.3. Escala logística	354
8.4. Estimação das Correlações Genéticas e Fenotípicas.....	354
8.5. Seleção e Ganho Genético na Escala Binomial	355
8.5.1. Seleção individual	355
8.5.2. Seleção pelo índice multiefeitos	356
8.5.3. Seleção de progênies e de genitores	356
8.5.4. Seleção com medidas repetidas.....	357
8.6. Análise Genética da Variável Sobrevida em Espécies Florestais	357
8.7. Estimação e Predição em Modelos Não Lineares Mistas Via Técnica de Modelos Lineares Generalizados	363
8.7.1. Introdução	363
8.7.2. Variáveis binomiais e funções de transformação	364
8.7.3. Modelos lineares generalizados	365
8.7.4. Estimação e predição em modelos não lineares.....	366
8.7.5. Modelos não lineares fixos	368
8.7.6. Modelos não lineares mistos	369
8.7.7. Aplicação a dados experimentais	373
8.7.7.1. Experimento 1 – Eucalipto.....	373
8.7.7.2. Experimento 2 – Erva-mate	376
8.7.7.3. Parâmetro de dispersão e ajuste da distribuição discreta aos dados.....	379

9. Estimação de Componentes de Variância via REML..... 384

9.1. Propriedades dos Estimadores.....	384
9.2. Principais Métodos de Estimação em Inferência Estatística Frequêntista	384
9.2.1. Quadrados mínimos (LS)	384
9.2.2. Máxima verossimilhança (ML)	385
9.2.3. Máxima verossimilhança restrita (REML)	385

9.2.4. Método R	387
9.3. Algoritmos para a Obtenção do Máximo da Função de Verossimilhança	388
9.4. Programas Computacionais	390
9.5. Estimadores ML e REML para Modelos com um Efeito Aleatório, Empregando o Algoritmo EM.....	391
9.6. Estimadores DF-REML.....	392
9.6.1. Estimadores DF-REML para modelos com um efeito aleatório.....	392
9.6.2. Estimadores DF-REML para modelos com mais de um efeito aleatório	394
9.7. Modelo Geral de Estimação REML e Predição BLUP	394
9.8. Estimação REML de Componentes de Variância e Predição BLUP de Valores Genéticos Intrapopulacionais	397
9.8.1. Delineamento em blocos ao acaso, progênies de polinização aberta, uma só população	397
9.8.1.1. 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	397
9.8.1.2. 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.....	400
9.8.1.3. 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 (ℓ) locais (experimentos) com algumas progênies ou tratamentos comuns	402
9.8.1.4. 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.....	405
9.8.2. Delineamento em blocos ao acaso, progênies de polinização controlada, uma só população	407
9.8.2.1. 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	407
9.8.2.2. 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.....	409
9.8.2.3. 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	410

9.8.2.4. 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	411
9.8.3. Delineamento em blocos ao acaso, progênies de polinização aberta, várias populações	412
9.8.3.1. 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	412
9.8.3.2. 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	416
9.8.3.3. 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	419
9.8.3.4. 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	422
9.8.4. Delineamento em látice, progênies de polinização aberta, uma população	424
9.8.4.1. 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	424
9.8.4.2. 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	426
9.8.4.3. 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 (ℓ) experimentos com algumas progênies ou tratamentos comuns	428
9.8.4.4. 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	430
9.8.5. Delineamento em látice, progênies de polinização aberta, várias populações	431

9.8.5.1. 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.....	431
9.8.5.2. 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	433
9.8.6. Avaliação apenas de populações	436
9.8.6.1. Avaliação de várias populações (procedências), no delineamento em blocos ao acaso, com várias plantas por parcela, uma medição por indivíduo e um só caráter	436
9.8.7. Avaliação simultânea de caracteres	438
9.8.7.1. 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	438
9.8.8. Espécies com sistema reprodutivo misto	440
9.9. Estimação REML de Componentes de Variância e Predição BLUP de Valores Genéticos Interpopulacionais	441
9.9.1. Híbridos envolvendo duas espécies ou populações	441
9.9.2. Híbridos envolvendo três ou mais espécies.....	451
9.10. Componentes de Variância e Valores Genotípicos em Testes Clonais.....	452
9.10.1. 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)	452
9.10.2. Avaliação de clones aparentados no delineamento em blocos ao acaso, com várias plantas por parcela.....	454
9.11. Escolha e Comparação entre Modelos de Análise	455
9.12. Estimação da Precisão das Estimativas dos Componentes de Variância	456
9.13. Exemplo Prático pelo DFREML	456
9.14. Estimação de Efeitos e Componentes de Variância Epistáticos	458
9.14.1. Efeitos genéticos e suas variâncias.....	458
9.14.2. Covariância entre parentes e estimação de variâncias epistáticas	459
9.14.3. Modelo linear misto incluindo efeitos epistáticos	462
9.14.4. Melhor predição linear não viciada (BLUP) dos efeitos epistáticos	463
9.14.5. Estimação dos componentes de variância por REML.....	464
9.14.6. Importância do ajuste dos efeitos de dominância e dos efeitos epistáticos do tipo aditivo x aditivo para a estimação de parâmetros genéticos e predição dos efeitos genéticos.....	465
9.15. Vantagens do Uso da Metodologia de Modelos Mistos para a Estimação de Componentes de Variância e Predição de Valores Genéticos.....	467
9.16. Modelos não Lineares para Variáveis Binomiais e Categóricas.....	468

9.17. Ajuste de Covariável e Análise de Covariância	473
---	-----

10. O Software Selegen – Seleção Genética Computadorizada 477

10.1. Introdução	478
10.2. Módulos Básicos	479
10.3. Entrada de Dados e Tratamento aos Valores Zero.....	479
10.4. Análises e Resultados Fornecidos pelo Módulo Índice de Seleção	480
10.5. Módulo BLP.....	488
10.6. Módulo BLUP	488
10.7. Módulo REML/BLUP: Modelos Computacionais.....	488
10.7.1. Modelos equivalentes	488
10.7.2. Modelos computacionais quando se conhecem ambos os genitores	490
10.7.2.1. Modelo individual – MI.....	490
10.7.2.2. Modelo inidividual reduzido – MIR	492
10.7.3. Modelos computacionais quando se conhece apenas um genitor	496
10.7.3.1. Modelo individual – MI.....	497
10.7.3.2. Modelo individual reduzido – MIR	499
10.7.3.3. Modelo gamético – MG	502
10.7.4. Modelo de clones repetidos – MCR.....	504

11. Regressão Aleatória e Funções de Covariâncias na Análise de Dados Longitudinais 507

11.1. Introdução	508
11.2. Regressão Aleatória no Contexto dos Modelos Lineares Mistos	509
11.3. Estimação dos Efeitos Fixos e Predição dos Efeitos Aleatórios.....	510
11.4. Estimação dos Componentes de Variância e Covariância.....	511
11.5. Predição dos Valores Genéticos em Cada Idade	512
11.6. Exemplo Numérico com Dados Simulados.....	512
11.7. Aplicação a Dados Experimentais	517
11.8. Síntese dos Métodos de Análise de Dados Longitudinais no Melhoramento	522

12. Inferência Estatística Bayesiana e Simulação Estocástica na Avaliação Genética 527

12.1. Sinopse	528
12.2. Introdução	531
12.3. Estimação e Predição Bayesiana	532
12.3.1. Conceito de probabilidade, probabilidade condicional e Teorema de Bayes.....	533
12.3.2. Função de verossimilhança	537
12.3.3. Distribuições a priori.....	540
12.3.4. Priori não informativa.....	540

12.3.5. Determinação subjetiva da priori	541
12.3.6. Determinação da priori por formas funcionais	541
12.3.7. Inferência bayesiana, inferência freqüentista e inferência verossimilhança	543
12.4. Fundamentos Bayesianos da Predição de Valores Genéticos.....	544
12.5. Estimação Bayesiana de Índices de Seleção Unicaracterísticos	545
12.6. Relação entre BLUP e Estimadores Bayesianos	546
12.7. Relação entre Estimadores de Máxima Verossimilhança e Estimadores Bayesianos.....	548
12.8. Estimação Bayesiana de Índices de Seleção Multicaracterísticos	548
12.9. Estimação Bayesiana de Componentes de Variância e Relação com ML e REML	549
12.10. Estimadores Melhorados para a Média de Tratamentos.....	551
12.11. Problemas da Abordagem Clássica de Predição de Valores Genéticos e Vantagens da Abordagem Bayesiana	553
12.12. Implementação Prática da Análise Bayesiana	559
12.13. Diagnóstico de Convergência.....	567
12.14. Estimação Bayesiana via Modelos Lineares Robustos.....	568
12.15. Aplicação a Dados Experimentais	570
13. Estatística Experimental: Comparação entre Tratamentos de Efeitos Fixos	577
13.1. Análise da Variância Paramétrica e Comparações Múltiplas	579
13.2. Alternativas aos Procedimentos de Comparações Múltiplas	585
13.3. Pressupostos da Análise de Variância e suas Verificações	586
13.3.1. Aditividade.....	586
13.3.2. Normalidade.....	586
13.3.3. Independência.....	588
13.3.4. Homocedasticidade.....	589
13.3.5. Transformações de dados	590
13.4. Análise de Variância Não Paramétrica	591
13.4.1. Análise de variância de Kruskal-Wallis para o delineamento inteiramente casualizado	591
13.4.2. Análise de variância de Friedman para o delineamento em blocos casualizados.....	593
13.5. Estimadores Melhorados para a Média de Tratamentos.....	595
13.6. Teste de Efeitos Fixos e Aleatórios no Contexto dos Modelos Mistos ...	600
13.6.1. Teste de efeitos fixos.....	600
13.6.2. Teste de efeitos aleatórios.....	603
13.7. Análise de Variância Multivariada.....	604
13.7.1. Teste T^2 de Hotelling	604
13.7.2. Teste Λ de Wilks (Manova)	606
13.8. Análise de Tabelas de Contingência	609

14. Delineamento de Experimentos para a Estimação de Parâmetros Genéticos e Seleção.....	615
14.1. Introdução	616
14.2. Minimização da Variância da Estimativa da Herdabilidade Baseada na Covariância de Meios-irmãos	616
14.3. Estimação da Herdabilidade via Regressão Pais-filhos	622
14.4. Estimação de Correlações Genéticas.....	623
14.5. Delineamento Baseado na Estimação Combinada via Covariância de Irmãos e Regressão Pais-filhos	624
14.6. Estimação da Repetibilidade	624
14.7. Delineamento para Maximização da Acurácia Seletiva	625
14.8. Repetibilidade e Número de Medições	627
14.9. Transformações de Herdabilidades	630
14.10. Delineamento para a Maximização da Acurácia Seletiva em Testes Clonais	631
14.11. Delineamentos para a Maximização da Acurácia Seletiva em Cruzamentos Fatoriais e Dialélicos.....	633
14.12. Tamanho da Parcela	633
14.13. Tamanho da População Experimental, Número de Famílias e Número de Repetições	638
14.14. Delineamentos Experimentais	638
14.15. Número Adequado de Locais de Experimentação para Um Número Total Fixo de Indivíduos a Serem Avaliados	643
14.16. Estimativas de Parâmetros e Controle Genético de Caracteres de Importância Econômica em Algumas Espécies Perenes.....	644
14.17. Obtenção de Estimativas Médias de Parâmetros Genéticos por Meio de Ponderações.....	655
14.18. Classificação das Magnitudes da Herdabilidade e da Acurácia Seletiva	656
14.19. Análise Espacial via REML/BLUP Individual.....	657
14.19.1. Introdução	657
14.19.2. Métodos de análise espacial e áreas de aplicação em genética.....	658
14.19.3. Modelos ARIMA na análise espacial de experimentos de campo	662
14.19.4. Modelos lineares mistos espaciais ao nível de indivíduos com estrutura auto-regressiva de erros	666
14.19.5. Aplicação a dados experimentais	672
15. Delineamentos de Cruzamento e Comparação de Métodos de Melhoramento.....	681
15.1. Introdução	682
15.2. Principais Delineamentos de Cruzamento	682
15.3. Comparação entre Delineamentos de Cruzamento	684
15.3.1. Modelos estatísticos	685

15.3.2. Índice multiefeitos para progênie de meios-irmãos e irmãos germanos	687
15.3.3. Índice multiefeitos associados a cruzamentos fatoriais.....	688
15.3.4. Parâmetros assumidos na comparação.....	690
15.3.5. Resultados e inferências práticas	691
15.3.6. Considerações finais	698
15.4. Comparação de Métodos de Melhoramento.....	699
15.5. Seleção de Espécies e Populações.....	708
15.6. Seleção de Cruzamentos Dentro de Populações	709
15.7. Cruzamentos Preferenciais	711
15.8. Delineamento de Cruzamento com Ênfase nos Melhores Indivíduos ...	714
16. Melhoramento de Híbridos Interpopulacionais e Seleção Recorrente em Espécies Perenes	717
16.1. Introdução	718
16.2. Estratégias de Melhoramento.....	718
16.3. Comparação entre Procedimentos de Seleção Recorrente Recíproca	722
16.4. Índice de Seleção para a Seleção Recorrente Recíproca.....	727
16.5. Seleção Recorrente em Algumas Espécies Perenes.....	728
16.5.1. Espécies florestais	729
16.5.2. Espécies produtoras de alimentos estimulantes	734
16.5.3. Palmáceas	736
16.5.4. Fruteiras.....	738
16.6. Esquema Detalhado para o Melhoramento de Híbridos Interpopulacionais	740
16.7. Melhoramento de Híbridos Triplos.....	744
16.8. Híbridos Interespecíficos Importantes no Gênero Eucalyptus.....	745
16.9. Heterozigoses e Heteroses Associadas a Híbridos Multiespécies	746
16.10. Estratégias de Seleção Recorrente em Função de Aspectos Biológicos das Espécies	750
17. Espécies Autógamas, Híbridos de Linhagens e Progênie S₁.....	753
17.1. Introdução	754
17.2. Espécies Autógamas Perenes.....	754
17.2.1. Modelos genéticos e componentes de variação genotípica	754
17.2.2. Covariância genética entre parentes	755
17.2.3. Comparação entre processos seletivos	756
17.2.4. Resultados comparativos entre métodos de melhoramento genético de espécies autógamas perenes	758
17.2.5. Estratégias de utilização do material genético melhorado	761
17.2.6. Estratégias para coleta de germoplasma, experimentação e melhoramento no longo prazo	762
17.3. Híbridos de Linhagens.....	763

17.3.1. Estimação de parâmetros genéticos a partir de cruzamentos entre várias linhagens	763
17.3.2. BLUP e predição de valores genotípicos por meio dos valores fenotípicos dos próprios híbridos	768
17.3.3. BLUP e predição de valores genotípicos de híbridos não avaliados	771
17.4. Uso de Progênies de Autofecundação em Espécies Alógamas	780
18. Espécies com Sistema Reprodutivo Misto	789
18.1. Introdução	790
18.2. Modelo Genético	790
18.3. Progresso Genético e Métodos de Seleção	792
18.4. Parâmetros Genéticos	795
18.5. Efeitos nas Estimativas de Parâmetros Genéticos e nos Métodos de Seleção	796
18.6. Comparação entre Tipos de Progênies para Maximização do Ganhos Genéticos	804
18.7. Tamanho Efetivo Populacional	806
19. Espécies de Reprodução Vegetativa e Testes Clonais	815
19.1. Introdução	816
19.2. Modelos Estatísticos Lineares	816
19.3. Análise de Variância, Esperança de Quadrados Médios e Testes dos Efeitos	818
19.4. Estimação dos Componentes da Variância Genética e Parâmetros Genéticos	820
19.5. Métodos de Seleção	823
19.6. Seleção a partir de Um Experimento ou Fase	824
19.7. Vantagens da Clonagem nos Testes de Progênies e Comparação com a Seleção Auxiliada por Marcadores (MAS)	827
19.8. Seleção a partir de Vários Experimentos Seqüenciais	829
19.9. Seleção de Clones Avaliados em Vários Espaçamentos ou Ambientes	835
19.10. Intereração Genótipo x Ambiente e Estratégias de Avaliação Genética e Seleção	838
19.11. Análise Ambiental a partir de Experimentos no Delineamento em Blocos	847
20. Limites Seletivos e Estruturas de Populações	853
20.1. Limites Seletivos	854
20.2. Seleção com Restrição ao Incremento da Endogamia	859
20.3. Estrutura de Populações de Melhoramento	860
20.3.1. Núcleo fechado de melhoramento	861
20.3.2. Núcleo aberto de melhoramento	862

20.3.3. Estrutura bidirecional em três níveis	862
20.3.4. Comparação das três estruturas no melhoramento de espécies perenes	864
20.3.5. Medidas de diversidade genética	867
21. Fundamentos de Genética de Populações para a Conservação de Germoplasma	869
21.1. Introdução	870
21.2. Genética de Populações e Conservação Biológica	870
21.3. Tamanho Efetivo Populacional e Deriva Genética.....	871
21.3.1. N_e e probabilidade de extinção	871
21.3.2. Estratégias para aumento do N_e e manutenção da variação genética	872
21.3.3. Proporção N_e/N em populações silvestres.....	873
21.3.4. Valores de N_e adequados à conservação genética	874
21.4. Endogamia e Depressão Endogâmica	875
21.4.1. Conceitos de endogamia e relações com a estrutura genética de populações e sistemas reprodutivos	875
21.4.2. Endogamia devida ao pequeno N_e	877
21.4.3. Efeito da endogamia na média populacional – depressão endogâmica	877
21.4.4. Efeito da endogamia na variância populacional – perda de variação genética.....	878
21.4.5. Depressão endogâmica em populações silvestres e populações florestais	878
21.5. Polimorfismo e Variação Genética.....	879
21.6. Mutação e Carga Genética.....	881
21.6.1. Carga genética	881
21.6.2. Mutação e manutenção da variabilidade genética	881
21.6.3. Acumulação e perda de mutações deletérias	883
21.7. Migração e Fluxo Gênico.....	883
21.8. Subdivisão e Estrutura de Populações	884
21.9. Estimação da Herdabilidade em Populações Naturais Empregando Dados Morfológicos e Marcadores Genéticos	887
21.10. Procedimentos <i>Bootstrap</i> e <i>Jackknife</i> no Estudo de Parâmetros em Genética de Populações	888
21.11. Seleção	890
21.12. Amostragem para Estudos em Genética de Populações	891
21.13. Estratégias para Implantação de Bancos de Germoplasma de Plantas Perenes	898
21.14. Recomendações	898
22. Referências Bibliográficas	901
Índice Remissivo de Espécies Perenes e Anuais	973

Notação Matemática e Estatística

Alfabeto grego

Letra	Nome	Letra	Nome	Letra	Nome
A α	alpha	I ι	lota	P ρ	rho
B β	beta	K κ	kappa	Σ σ	sigma
Γ γ	gamma	Λ λ	lamba	T τ	tau
Δ δ	delta	M μ	mu (mi)	Υ υ	upsilon
E ϵ	épsilon	N ν	nu (ni)	Φ ϕ φ	phi
Z ζ	zeta	Ξ ξ	xi (qsi)	X χ	chi (qui)
H η	eta	O \circ	omicron	Ψ ψ	psi
Θ θ ϑ	theta	Π π	pi	Ω ω	ômega

Símbolos Matemáticos

- \int : operador de integração (integral de);
 ∂ : operador de derivação parcial de funções de duas ou mais variáveis (del);
 $\lim_{x \rightarrow \infty} f(x)$: limite de uma função $f(x)$ quando x tende ao infinito;
 Σ : somatório;
 \prod : produtório;
 \cup : união ou soma lógica;
 \cap : interseção ou produto lógico;
 ∞ : infinito;
 \rightarrow : tende a;
 \propto : proporcional a;
 \equiv : identidade ou idêntico a;
 $=$: igual a;
 \approx ou \sim : aproximadamente igual a;
 \neq : diferente;
 $<$: menor que;
 $>$: maior que;
 \leq : menor ou igual a;
 \geq : maior ou igual a;
 \in : pertence a;
 \notin : não pertence a;
 \forall : todo e qualquer;
 $|x|$: módulo ou valor absoluto de x ;
 $\#$: cardinal : símbolo de contagem que indica número de elementos;
 θ_i^k : valor do parâmetro θ_i na iteração k.

Simbologia Estatística

- \sim : distribuída como;

$E(y)$: esperança matemática da variável aleatória y ;
 $\text{Var}(y)$: variância da variável aleatória y ;
 $\text{Cov}(x,y)$: covariância entre as variáveis aleatórias x e y ;
 \hat{a} : estimador ou estimativa do parâmetro a ;
 $\theta|y$: distribuição condicional de θ dado y ;
 $p(\theta|y)$: probabilidade condicional de θ dado y ;
 Ω : espaço amostral;
 $\ell(\theta; y)$: função de verossimilhança de θ , a qual associa a cada valor de θ , o valor $f(y|\theta)$;
 α : nível de significância (probabilidade de erro tipo I);
 β : probabilidade de erro tipo II;
 $1-\alpha$: grau ou coeficiente de confiança;
 $1-\beta$: poder de um teste estatístico;
 χ^2 : estatística com distribuição qui-quadrado;

Operadores Matriciais

\otimes : produto de Kronecker;
 \oplus : soma direta;
 $\#$: produto de Hadamard;
 tr : operador traço matricial;
 $r(X)$: posto da matriz X ;
 A^{-1} : inversa comum da matriz A ;
 $A^{-\cdot}$: inversa generalizada da matriz A ;
 A' : transposta da matriz A ;

Notação de alguns parâmetros genéticos e fenotípicos

y : valor fenotípico, observação ou vetor de observações;
 a : efeito genético aditivo ou vetor de efeitos genéticos aditivos;
 d : efeito genético de dominância ou vetor de efeitos de dominância;
 g : efeito genotípico ($a + d$) ou vetor de efeitos genotípicos;
 δ : efeito de indivíduo dentro de parcela;
 β : vetor de efeitos fixos;
 σ_y^2 : variância fenotípica individual;
 σ_a^2 : variância genética aditiva;
 σ_d^2 : variância genética de dominância;
 h_a^2 : herdabilidade individual no sentido restrito ou herdabilidade dos efeitos aditivos;
 h_g^2 : herdabilidade individual no sentido amplo ou herdabilidade dos efeitos genotípicos;
 c^2 : correlação intraclasse devida ao ambiente comum da parcela;
 p^2 : coeficiente de determinação dos efeitos de ambiente permanente (confundidos com os efeitos de dominância em alguns modelos);
 ρ : coeficiente de repetibilidade individual;

ρ_a : correlação genética aditiva intraclasse entre os indivíduos da progênie;

ρ_d : correlação genética de dominância intraclasse entre os indivíduos da progênie;

ρ_p : correlação genética intraclasse entre os indivíduos de diferentes progênies, mas da mesma população;

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

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

A#A: matriz de correlação genética epistática (do tipo aditivo x aditivo) entre os indivíduos em avaliação;

A#D: matriz de correlação genética epistática (do tipo aditivo x dominante) entre os indivíduos em avaliação;

D#D: matriz de correlação genética epistática (do tipo dominante x dominante) entre os indivíduos em avaliação;

A#A#A: matriz de correlação genética epistática (do tipo aditivo x aditivo x aditivo) entre os indivíduos em avaliação;

D₁: covariância entre os efeitos aditivos dos alelos e os efeitos de dominância dos homozigotos;

D₂: variância genética dos efeitos de dominância dos homozigotos;

H: soma dos quadrados da depressão por endogamia de cada loco;

H²: quadrado da soma da depressão por endogamia;

σ_i^2 : variância do efeito i;

$\lambda = \frac{\sigma_e^2}{\sigma_i^2}$: relação entre a variância residual e a variância do efeito i;

t₁: correlação intraclasse entre os indivíduos de uma mesma parcela;

t₂: correlação intraclasse entre os indivíduos de uma mesma família em diferentes blocos;

t₃: correlação intraclasse entre os indivíduos de diferentes famílias, mas da mesma população;

r_{aa} : acurácia seletiva ou da predição dos valores genéticos aditivos;

r_{gg} : acurácia seletiva ou da predição dos valores genotípicos;

r_{gg} : correlação genética entre o desempenho do germoplasma através dos ambientes;

PEV: variância do erro de predição dos valores genéticos;

N_e: tamanho efetivo populacional;

N_{ef}: número efetivo de famílias selecionadas;

n: número de plantas por parcela;

b: número de repetições;

F: coeficiente de endogamia;

r_a: correlação genética entre caracteres;

r_{x,y} ou r: correlação fenotípica entre caracteres;

r_{fp}: correlação fenotípica permanente entre caracteres;

I: índice de seleção ou índice de predição dos valores genéticos;

G_s: ganho genético com a seleção;

h_c^2 : coherdabilidade ou herdabilidade indireta ou correlacionada;

$\psi_{(x,y)}$: coeficiente de predição genética.

ψ_m : coeficiente de predição genética multivariado;

h_m^2 : herdabilidade mutivariada;

β' : herdabilidade generalizada.

Siglas

BLUP: Melhor Preditor Linear não Viciado;

EBLUP: Melhor Preditor Linear não Viciado Empírico;

IME: Índice multiefeitos;

BLP: Melhor Preditor Linear;

BP: Melhor Preditor;

MAP: Máximo a Posteriori;

EB: Estimador bayesiano;

EMMSE: Estimador de Mínimo Erro Quadrático Médio;

REML: Máxima Verossimilhança Restrita;

ML: Máxima Verossimilhança;

VEIL: Método de Estimação de Componentes de Variância por Verossimilhança Integrada;

\mathcal{R} : Método \mathcal{R} (Regressão) de Estimação de Componentes de Variância;

LS: Quadrados Mínimos;

OLS: Quadrados Mínimos Ordinários;

GLS: Quadrados Mínimos Generalizados;

BLUE: Melhor Estimador Linear não Viciado;

BUE: Melhor Estimador não Viciado;

BQUE: Melhor Estimador Quadrático não Viciado;

EM: Algoritmo *Expectation- Maximization*;

DF: Algoritmo não Derivativo;

AI: Algoritmo de Informação Média;

EM-REML: REML obtido via algoritmo EM;

DF-REML: REML obtido via algoritmo DF;

AI-REML: REML obtido via algoritmo AI;

GLM : Modelos Lineares Generalizados;

EMM : Equações de Modelo Misto;

PEV : Variância do Erro de Predição dos Valores Genéticos;

Bootstrap: Método de Reamostragem com Reposição;

Jackknife: Método de Reamostragem sem Reposição;

CGC: Capacidade Geral de Combinação;

CEC: Capacidade Específica de Combinação;

CGH: Capacidade Geral de Hibridação;

MI: Progênes de Meios-irmãos;

IG: Progênes de Irmãos Germanos;

QTL: Locus de Características Quantitativas;

MAS: Seleção Auxiliada por Marcadores;

CGS: Seleção de Genes Candidatos.

Capítulo Introdutório

**Espécies Perenes,
Sistemas Reprodutivos
e Metodologias de Seleção**

As espécies vegetais perenes, como as essências florestais, as palmáceas e as frutíferas arbóreas, apresentam vários aspectos biológicos peculiares, que tornam o seu melhoramento genético bastante diferenciado do melhoramento de culturas anuais. Dentre os referidos aspectos, citam-se: sobreposição de gerações, ciclo reprodutivo longo, reprodução sexuada e assexuada, expressão dos caracteres ao longo das várias idades.

Em consequência destes fatores, verificam-se os seguintes reflexos no melhoramento destas espécies: utilização dos indivíduos selecionados, para produção durante vários anos, fato que demanda muito rigor e precisão nos métodos de seleção; uso de avaliações repetidas em cada indivíduo ao longo do tempo; seleção envolvendo comparações de indivíduos de diferentes gerações e, portanto, avaliados em diferentes condições ambientais, fato que requer o uso de métodos de avaliação genética mais elaborados; seleção também para os efeitos não aditivos (além da seleção para os efeitos aditivos) dos alelos tendo em vista a propagação vegetativa dos indivíduos selecionados; relevância da unidade de seleção “indivíduo” em detrimento da unidade de seleção “média de grupos de indivíduos”, fato que demanda a predição dos valores genéticos (aditivos e não aditivos) individuais para fins de seleção; redução na taxa de sobrevivência das plantas nos experimentos ao longo das idades, fato que, associado à sobreposição de gerações, tende a gerar dados desbalanceados para uso na estimativa de parâmetros genéticos e na predição dos valores genéticos individuais. Outro fator a considerar nestas espécies é a possível alteração no controle genético dos caracteres ao longo das idades.

A maioria das espécies perenes é alógama (menos de 5% de autofecundação) ou apresentam sistema reprodutivo misto (em que predominam os cruzamentos, mas ocorrem autofecundações em taxas inferiores a 50% e superiores a 5%), sendo raras as autógamas. A maioria das espécies perenes permite, também, a propagação vegetativa. Na Tabela 1 são apresentadas algumas das principais espécies perenes cultivadas no Brasil e seus sistemas reprodutivos.

O conceito de espécies perenes empregado neste documento refere-se ao tradicional, baseado na possibilidade de os indivíduos se reproduzirem mais de uma vez e/ou permitir mais de uma avaliação do caráter no decorrer de suas vidas. Dessa forma, engloba, também, sem restrição, todas as espécies de propagação vegetativa, as quais são essencialmente perenes (Simmonds, 1979), uma vez que todos os clones são potencialmente imortais e, portanto, permitem a avaliação de um caráter seguidas vezes no indivíduo, bem como a utilização comercial do mesmo indivíduo inúmeras vezes.

Para o melhoramento de todas as espécies apresentadas na Tabela 1 e outras perenes não relatadas, o presente documento é potencialmente impor-

tante para o aumento da eficiência do processo seletivo. Dentre os grupos de espécies que mais se beneficiam das metodologias biométricas abordadas destacam-se: (i) espécies florestais (como o eucalipto, o pinus, a acácia-negra, a seringueira, a leucena, a grevílea, a teca, o mogno); (ii) espécies produtoras de alimentos estimulantes contendo os alcalóides cafeína e teobromina (como o café, o cacau, a erva-mate, o guaraná, o chá-da-índia); (iii) palmáceas, produtoras de palmitos, óleos e frutos comestíveis (como o dendê, o coco, a pupunha, o açaí, a juçara, a palmeira-real, a tamareira); (iv) fruteiras (como a acerola, o caju, o cupuaçu, a maçã, a graviola, a pinha); (v) espécies forrageiras (como o capim elefante, a braquiária, o panicum, os estilosantes, a alfafa); (vi) espécies energéticas (como a cana-de-açúcar, a mandioca); (vii) espécies hortícolas ou olerícolas (como a batata, o morango).

Tabela 1. Principais espécies perenes cultivadas no Brasil e seus sistemas reprodutivos.

Espécies	Nome científico	Sistema reprodutivo*
<u>Espécies Florestais (Madeireiras)</u>		
Eucalipto	<i>Eucalyptus</i> spp.	Misto/Alogamia
Pinus	<i>Pinus</i> spp.	Alogamia
Acácia-negra	<i>Acacia mearnsii</i>	Alogamia
Leucena	<i>Leucaena leucocephala</i>	Autogamia
Grevilea	<i>Grevillea robusta</i>	Alogamia
Algaroba	<i>Prosopis juliflora</i>	Alogamia
Teca	<i>Tectona grandis</i>	Alogamia
Gmelina	<i>Gmelina arborea</i>	Alogamia
Acacia-mangium	<i>Acacia mangium</i>	Alogamia
Acacia-auriculiformis	<i>Acacia auriculiformis</i>	Alogamia
Bracatinga	<i>Mimosa scabrella</i>	Alogamia
Pinheiro-do-paraná	<i>Araucaria angustifolia</i>	Alogamia (Dióica)
Castanha-do-pará	<i>Bertholetia excelsa</i>	Misto/Alogamia
Morototó	<i>Didymopanax morototoni</i>	Alogamia
Taxi-branco	<i>Sclerobium paniculatum</i>	Alogamia
Freijó-cinza	<i>Cordia goeldiana</i>	Alogamia
Mogno	<i>Swietenia macrophylla</i>	Alogamia
Mogno Africano	<i>Khaya ivorensis</i>	Alogamia
Nim Indiano	<i>Azadirachta indica</i>	Alogamia
<u>Espécies Florestais (Extrativas)</u>		
Seringueira	<i>Hevea brasiliensis</i>	Misto/Alogamia
Erva-mate	<i>Ilex paraguariensis</i>	Alogamia (Dióica)
Chá-da-índia	<i>Camellia sinensis</i>	Alogamia

Continua...

Espécies	Nome científico	Sistema reprodutivo*
<u>Espécies Florestais (Extrativas)</u>		
Palmiteiro-juçara	<i>Euterpe edulis</i>	Alogamia
Palmiteiro-açaí	<i>Euterpe oleraceae</i>	Alogamia
Pupunha	<i>Bactris gasipaes</i>	Alogamia
Palmeira-real	<i>Archontophoenix spp.</i>	Alogamia/Misto
Pimenta-longa	<i>Piper hispidinervum</i>	Alogamia
Cravo-da-índia	<i>Caryophyllus aromaticus</i>	Alogamia
Canela-da-índia	<i>Cinnamomum zeylanicum</i>	Alogamia
<u>Espécies Frutíferas</u>		
Abacate	<i>Persea americana</i>	Alogamia
Abacaxi	<i>Ananas comosus</i>	Assexual (Vegetativa)
Açaí**	<i>Euterpe oleraceae</i>	Alogamia
Acerola	<i>Malpighia emarginata</i>	Alogamia
Banana	<i>Musa spp.</i>	Assexual (Vegetativa)
Caqui	<i>Diospyrus ebenaster</i>	Alogamia
Cupuaçu	<i>Theobroma grandiflorum</i>	Alogamia
Figo	<i>Ficus carica</i>	Alogamia
Goiaba	<i>Psidium guajava</i>	Alogamia/Misto
Graviola	<i>Annona muricata</i>	Alogamia
Kiwi	<i>Actinidia chinensis</i>	Alogamia (Dióica)
Laranja	<i>Citrus spp.</i>	Assexual (Poliembrionia nucelar)
Limão	<i>Citrus spp.</i>	Assexual (Poliembrionia nucelar)
Maçã	<i>Malus domestica</i>	Alogamia
Mamão	<i>Carica papaya</i>	Alogamia
Manga	<i>Mangifera indica</i>	Alogamia
Maracujá	<i>Passiflora edulis</i>	Alogamia
Nogueira-macadâmia	<i>Macadamia integrifolia</i>	Alogamia
Nogueira-pecã	<i>Carya illinoensis</i>	Alogamia/Misto
Pêra	<i>Pyrus spp.</i>	Alogamia
Pêssego	<i>Prunus persica</i>	Autogamia
Pinha	<i>Annona squamosa</i>	Alogamia/Misto
Pequi	<i>Caryocar brasiliense</i>	Alogamia/Misto
Pupunha**	<i>Bactris gasipaes</i>	Alogamia
Tamareira	<i>Phoenix dactylifera</i>	Alogamia (Dióica)
Umbu	<i>Spondias tuberosa</i>	Alogamia
Uva	<i>Vitis vinifera</i>	Alogamia

Continua...

Espécies	Nome científico	Sistema reprodutivo*
Espécies Frutíferas (Agroindustriais)		
Cacaueiro	<i>Theobroma cacao</i>	Alogamia/Misto
Cafeeiro-arábica	<i>Coffea arabica</i>	Autogamia
Cafeeiro-robusta	<i>Coffea canephora</i>	Alogamia
Cajueiro	<i>Anacardium occidentale</i>	Misto/Alogamia
Coqueiro	<i>Cocos nucifera</i>	Alogamia/ Misto/Autogamia
Dendezeiro	<i>Elaeis guineensis</i>	Alogamia
Dendezeiro-caiauá	<i>Elaeis oleifera</i>	Alogamia
Guaranazeiro	<i>Paullinia cupana</i>	Alogamia
Pimenta-do-reino	<i>Piper nigrum</i>	Assexual (Vegetativa)
Urucum	<i>Bixa orellana</i>	Alogamia
Babaçu	<i>Attalea speciosa</i>	Alogamia
Carnaúba	<i>Copernicia prunifera</i>	Alogamia
Oliveira	<i>Olea europaea</i>	Alogamia
Espécies Forrageiras e Energéticas		
Alfafa	<i>Medicago sativa</i>	Alogamia/Assexual
Algaroba***	<i>Prosopis juliflora</i>	Alogamia
Braquiária	<i>Brachiaria spp.</i>	Assexual (Apomixia)
Cana-de-açúcar	<i>Saccharum spp.</i>	Assexual/Alogamia
Capim colonião	<i>Panicum maximum</i>	Assexual (Apomixia)
Capim-elefante (Napier)	<i>Pennisetum purpureum</i>	Assexual (Vegetativa)
Estilosantes	<i>Stylosanthes guianensis</i>	Autogamia
Leucena***	<i>Leucaena leucocephala</i>	Autogamia
Mandioca	<i>Manihot esculenta</i>	Assexual (Vegetativa)
Sabiá	<i>Mimosa caesalpiniaefolia</i>	Alogamia
Jurema-preta	<i>Mimosa hostilis</i>	Alogamia
Palma-forrageira	<i>Opuntia ficus-indica</i>	Assexual (Vegetativa)
Espécies Olerícolas		
Alho	<i>Allium sativum</i>	Assexual (Vegetativa)
Batata	<i>Solanum tuberosum</i>	Assexual (Vegetativa)
Batata-doce	<i>Ipomoea batatas</i>	Assexual (Vegetativa)
Morango	<i>Fragaria spp.</i>	Assexual (Vegetativa)
Gengibre	<i>Zingiber officinale</i>	Assexual (Vegetativa)
Inhame	<i>Colocasia esculenta</i>	Assexual (Vegetativa)
Espécies fibrosas		
Algodão-arbóreo	<i>Gossypium barbadense</i>	Misto
Amoreira	<i>Morus alba</i>	Assexual (Vegetativa)
Rami	<i>Bohemeria spp.</i>	Assexual/Alogamia
Sisal	<i>Agave sisalana</i>	Assexual (Vegetativa)

* Fontes: Simmonds (1979); Borém (1997); Resende (1999a).

** Espécies também incluídas como florestal extrativa, devido à produção de palmito;

*** Espécies também incluídas como florestais madeireiras.

O sucesso de um programa prático de melhoramento genético de espécies perenes depende, fundamentalmente, de conhecimentos sólidos em pelo menos quatro áreas: (i) **conhecimento do produto final de interesse**, com suas relações industriais e mercadológicas, exigências qualitativas e formas de uso pelo consumidor; (ii) **conhecimento do germoplasma disponível para a obtenção de tal produto**, notadamente da variação biológica entre espécies no gênero, entre populações dentro de espécie e dentro de populações; (iii) **conhecimento acerca de metodologias de seleção e de melhoramento**, destacando-se o emprego das técnicas de genética quantitativa; (iv) **conhecimento profundo dos fatores ambientais que afetam a expressão fenotípica**, notadamente os fatores edáficos, climáticos e técnicas de cultivo das plantas e de colheita dos produtos. O presente documento pretende contribuir para o aumento da eficiência dos programas de melhoramento genético de plantas perenes, sobretudo para os fatores relatados em (ii) e (iii).

No âmbito mundial, há carência de livros sobre melhoramento de plantas perenes, principalmente em termos de uma abordagem adequada do ponto de vista da genética quantitativa. Neste livro, os efeitos genéticos são tratados como variáveis aleatórias, na maioria das situações. Este fato remete os desenvolvimentos metodológicos à técnica de modelos lineares e não lineares mistos, em nível de indivíduos, como abordagem conceitual e ferramenta prática padrões para a predição de valores genéticos e estimação de componentes de variância. Tal abordagem vem sendo aplicada aos programas operacionais de melhoramento de plantas perenes no Brasil desde 1993 (Resende et al., 1993; Resende & Higa, 1994a; Resende et al., 1996b; Bueno Filho, 1997; Resende, 1999b; Bueno Filho & Vencovsky, 2000; Resende & Dias, 2000, dentre vários outros) e constitui-se em uma importante contribuição ao melhoramento destas plantas.

De maneira genérica, a metodologia de modelos mistos envolve a predição de valores genéticos pelo procedimento da melhor predição linear não viciada (BLUP) e a estimação de componentes de variância pelo método da máxima verossimilhança restrita (REML). O procedimento BLUP foi desenvolvido e aperfeiçoado pelo cientista Charles Roy Henderson (1911-1989) enquanto pesquisador na Universidade de Cornell-USA, conforme registrado em seus principais trabalhos (Henderson et al., 1959; Henderson, 1963; 1973; 1975; 1976; 1984). O método REML foi desenvolvido e melhorado pelo pesquisador Robin Thompson (atualmente chefe do Departamento de Matemática Biométrica e Estatística da *Rothamsted Experimental Station* – Inglaterra) e co-autores. Tal método (Patterson & Thompson, 1971) surgiu a partir de esforços na obtenção de melhores estimadores de componentes de variância para dados não ortogonais e desbalanceados (Thompson, 1969). Posteriormente, foi estendido para modelos multivariados (Thompson, 1973) e, recentemente, melhorado em termos do algoritmo de estimação via informação média (AI-REML) (Johnson & Thompson, 1995), visando a incorporação em softwares de excelência como o

GENSTATe o ASREML (Gilmour & Thompson, 1998). Os princípios gerais destes métodos de predição e estimação foram desenvolvidos. Todavia, as derivações de preditores e estimadores específicos associados aos vários modelos estatísticos e situações práticas na experimentação e melhoramento de várias categorias de plantas e animais foram realizadas por inúmeros autores.

É importante relatar que, para o caso balanceado, o procedimento BLUP individual é equivalente ao método do índice multiefeitos, o qual engloba todos os efeitos aleatórios do modelo estatístico. Neste mesmo caso, os métodos REML e análise de variância produzem estimativas idênticas para os componentes de variância.

Recentemente, a metodologia de modelos “mistos” tem sido abordada e aplicada também sob o enfoque Bayesiano. Em 1986, o trabalho clássico de Gianola & Fernando (1986) propôs a abordagem Bayesiana como uma estratégia conceitual para resolução de problemas em melhoramento genético. No Brasil, esta abordagem vem sendo recomendada e aplicada na avaliação genética em melhoramento animal (Magnabosco, 1997; Rosa, 1999) e de plantas perenes (Resende, 1997; 1999b). É relevante ressaltar que, no contexto Bayesiano não existem efeitos fixos mas somente variáveis aleatórias, ou seja, todos os parâmetros de interesse são variáveis aleatórias. Assim, modelo “misto” neste caso, refere-se a modelos com efeitos genéticos e com efeitos ambientais identificáveis (efeitos fixos, no contexto frequentista), ambos tratados como variáveis aleatórias. Tais variáveis são estimadas, diferentemente da abordagem frequentista, em que os efeitos aleatórios são preditos e os efeitos fixos e componentes de variância são estimados.

Previamente aos trabalhos de Henderson e de Thompson, vários trabalhos e autores foram fundamentais para o próprio desenvolvimento das ciências da Estatística e da Genética Quantitativa, bem como dos procedimentos de avaliação genética. O surgimento da Estatística deveu-se muito aos trabalhos de Karl Pearson (1857-1936) e Ronald Aylmer Fisher (1890-1962). Pearson (1903) derivou as médias e variâncias condicionais para a distribuição normal multivariada. Os resultados de Pearson foram apresentados em notação matricial por Aitken (1934) e empregados por Henderson no contexto dos preditores BLUP, os quais podem ser vistos como valores genéticos condicionais a um conjunto de ($N-r$) funções lineares dos dados, linearmente independentes e invariantes à translação, em que N é o número de observações e r é o posto de X , a matriz de incidência para os efeitos fixos. Os índices de seleção podem ser vistos como computações das médias condicionais dos valores genéticos dado as observações. Lush (1931) foi o primeiro cientista a utilizar preditores de valores genéticos baseados em médias condicionais e Cochran (1951) estendeu as propriedades ótimas dos índices de seleção para quaisquer distribuições.

Fisher desenvolveu simultaneamente a Estatística (ambas, Teórica e Experimental) e a Genética Quantitativa. Em 1918, publicou trabalho fundamental sobre componentes da variância genética e medidas de parentesco genético aditivo, de dominância e epistática (Fisher, 1918). Em 1922 criou o método da

máxima verossimilhança (Fisher, 1922) e, em 1925, desenvolveu a análise de variância, com implicações na estimativa de componentes de variância e delineamentos experimentais (Fisher, 1925). A abordagem inicial da análise de dados desbalanceados é devida a Fisher e Yates, ambos trabalhando na *Rothamsted Experimental Station*. Para este caso de representação desbalanceada, Yates (1934) apresentou as soluções de quadrados mínimos para dois diferentes modelos de classificação cruzada. Pela abordagem de Yates os valores genéticos eram estimados como efeitos fixos.

A Estatística é uma ciência nova, um produto do século XX. O matemático Karl Pearson fundou a primeira escola de Estatística e também o periódico *Biometrika* na Inglaterra. A aplicação da Estatística e Matemática a problemas biológicos deu origem à ciência da Biometria, Estatística Biométrica ou Matemática Biométrica. A união da teoria matemática, estatística e genética contribuiu para formular a ciência denominada Genética Quantitativa, Genética Biométrica, Genética Estatística ou Genética Matemática. Esta ciência é fundamental ao melhoramento de plantas e animais.

Capítulo 1

**Conceitos Básicos
de Genética Quantitativa
e Parâmetros Genéticos**

1.1. Base Genética dos Caracteres Quantitativos

Caracteres qualitativos são aqueles governados por um pequeno número de genes, que apresentam pequeno número de genótipos e distribuição descontínua de fenótipos, de forma que a cada fenótipo podem ser associados um ou poucos genótipos. O estudo desses caracteres é bastante simplificado e pode ser realizado por meio da Genética Qualitativa ou Mendeliana.

Entretanto, a maioria dos caracteres de interesse econômico apresenta herança quantitativa e devem ser utilizados procedimentos de Genética Quantitativa para analisá-los e estudá-los. Os caracteres quantitativos podem ser subdivididos em caracteres que apresentam distribuição contínua (volume de madeira, peso de frutos, peso de sementes, massa foliar, etc.) e descontínua (número de frutos, número de sementes por fruto, número de perfilhos, sobrevivência, etc.) de fenótipos, caracterizando-se como variáveis aleatórias contínuas e discretas, respectivamente. Os caracteres quantitativos com distribuição discreta de fenótipos são denominados características de limiar (Falconer, 1989), são governados por um grande número de genes, apresentam grande número de genótipos e elevadas influências ambientais, fatos que os caracterizam como quantitativos.

Os caracteres quantitativos são regidos por um grande número de genes (o que conduz a um grande número de genótipos) e/ou apresentam elevada influência ambiental. Estes fatores conduzem a um grande número de fenótipos de forma que não é possível associar genótipos aos diferentes fenótipos. Em caracteres quantitativos, genótipos iguais podem apresentar fenótipos diferentes bem como genótipos diferentes podem apresentar fenótipos similares, como resultados dos efeitos ambientais.

1.2. Modelo Genético e Fenotípico

Para populações panmíticas, o modelo genético geral para estudo dos caracteres quantitativos, ignorando a epistasia, equivale a:

$g = \mu + a + d$, em que:

g : valor genotípico do indivíduo, para um dado caráter.

μ : média genotípica do caráter na população.

a : efeito aditivo dos alelos do indivíduo.

d : efeito ou desvio de dominância.

$\mu + a$: valor genético aditivo ou valor genético.

O modelo fenotípico para estudo dos caracteres quantitativos é:

$y = \mu + a + d + e_p + e_t$, em que:

y : valor fenotípico do indivíduo.

e_p : efeito de ambiente permanente.

e_t : efeito de ambiente temporário.

$\mu + a + d + e_p$: valor fenotípico permanente.

Para caracteres avaliados uma única vez no indivíduo, tem-se $e = e_p + e_t$, portanto (Turner & Young, 1969; Falconer, 1989):

$$y = \mu + a + d + e = g + e$$

Quando indivíduos são avaliados em parcelas experimentais com várias plantas e são realizadas várias medições por indivíduo, tem-se: $y = \mu + a + d + c + e_p + e_t$, em que c refere-se ao efeito de ambiente comum da parcela.

Segundo esses modelos são os seguintes os valores esperados:

$$E(a) = E(d) = E(e) = E(e_t) = E(e_p) = E(c) = 0$$

$$E(y) = \mu$$

1.3. Componentes da Variação Fenotípica

Para um caráter avaliado em um único ambiente e em uma população panmítica são os seguintes os componentes de variação genética e fenotípica:

$\sigma_g^2 = \sigma_a^2 + \sigma_d^2$: variância genotípica.

$\sigma_y^2 = \sigma_a^2 + \sigma_d^2 + \sigma_e^2$: variância fenotípica.

Estes resultados são válidos considerando a ausência de correlação entre efeitos genéticos e ambientais. Esta ausência de correlação é conseguida por meio da casualização dos genótipos nos ambientes. Uma correlação entre genótipo e ambiente poderia resultar de melhor tratamento (ambiente) dado aos melhores genótipos.

Para caracteres avaliados mais de uma vez no indivíduo, supondo também ausência de correlação entre efeitos genéticos e ambientais, tem-se:

$\sigma_y^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2 + \sigma_{et}^2$: variância fenotípica.

As expressões apresentadas para a variância fenotípica incluem o componente σ_c^2 , quando parcelas com várias plantas são empregadas.

Definem-se os seguintes componentes de variância:

σ_g^2 : variância genotípica.

σ_y^2 : variância fenotípica.

σ_a^2 : variância genética aditiva.

σ_d^2 : variância genética de dominância.

σ_{ep}^2 : variância ambiental permanente.

σ_{et}^2 : variância ambiental temporária.

σ_e^2 : variância ambiental.

σ_c^2 : variância entre parcelas.

1.4. Covariância Genética entre Parentes

Considerando-se os indivíduos x e y, tem-se a seguinte expressão geral para a covariância genética entre parentes em uma população panmítica:

$$COV_g(x, y) = 2r_{xy}\sigma_a^2 + \mu_{xy}\sigma_d^2, \text{ em que:}$$

r_{XY} = coeficiente de parentesco de Malecot ou coeficiente de coancestria entre os indivíduos x e y.

μ_{XY} = coeficiente de parentesco de dominância entre os indivíduos x e y.

Com base nesta expressão geral, obtêm-se as seguintes covariâncias genéticas entre os parentes:

(1) Covariância entre mãe (x) e filho ou filha (y).

$$COV_g(M, F) = 2r_{XY}\sigma_a^2 = (1/2)\sigma_a^2$$

(2) Covariância entre pai (x) e filho ou filha (y).

$$COV_g(P, F) = 2r_{xy}\sigma_a^2 = (1/2)\sigma_a^2$$

(3) Covariância entre meios-irmãos x e y.

$$COV_g(MI) = 2r_{xy}\sigma_a^2 = (1/4)\sigma_a^2$$

(4) Covariância entre o indivíduo (x) e ele mesmo (variância entre indivíduos).

$$COV_g(I, I) = 2r_{xy}\sigma_a^2 + \mu_{xy}\sigma_d^2 = \sigma_g^2 = \sigma_a^2 + \sigma_d^2$$

(5) Covariância genética entre irmãos germanos x e y.

$$COV_g(IG) = 2r_{xy}\sigma_a^2 + \mu_{xy}\sigma_d^2 = (1/2)\sigma_a^2 + (1/4)\sigma_d^2$$

1.5. Correlação Genética entre Parentes e Correlação Fenotípica Intraclasse

A correlação genética aditiva intraclasse ou correlação genética entre dois indivíduos x e y é dada por:

$$\rho_a = 2r_{xy} = COV_a / (\sigma_a \sigma_a)$$

Assim, tem-se:

$$\rho_a = \frac{(1/4)\sigma_a^2}{\sigma_a \sigma_a} = 1/4 \text{ para meios-irmãos.}$$

$$\rho_a = \frac{(1/2)\sigma_a^2}{\sigma_a \sigma_a} = 1/2 \text{ para irmãos germanos.}$$

A correlação genética de dominância intraclasse entre os parentes x e y é dada por $\mu_{xy} = \rho_d = 0$ para meios-irmãos e $\mu_{xy} = \rho_d = 1/4$ para irmãos germanos.

A correlação fenotípica intraclasse entre dois indivíduos é dada por:

$$t = \frac{\text{cov}_y}{\sigma_y \sigma_y}, \text{ em que cov}_y \text{ é a covariância fenotípica entre os indivíduos.}$$

Dessa forma:

$$t = \frac{\rho_a \sigma_a^2 + \rho_d \sigma_d^2 + \sigma_c^2}{\sigma_y^2} = \rho_a h_a^2 + \rho_d (h_g^2 - h_a^2) + c^2$$

$$t = \frac{(1/4) \sigma_a^2 + \sigma_c^2}{\sigma_y^2} = (1/4) h_a^2 + c^2 \text{ para meios-irmãos.}$$

$$t = \frac{(1/2) \sigma_a^2 + (1/4) \sigma_d^2 + \sigma_c^2}{\sigma_y^2} = (1/2) h_a^2 + (1/4) (h_g^2 - h_a^2) + c^2 \text{ para irmãos germanos, em que:}$$

h_a^2 : herdabilidade individual no sentido restrito ou herdabilidade dos efeitos aditivos.

σ_c^2 : variância de ambiente comum ou variância entre parcelas quando os indivíduos em consideração encontram-se na mesma parcela.

$c^2 = \sigma_c^2 / \sigma_y^2$: fração da correlação intraclasse devida à correlação de ambiente comum.

h_g^2 : herdabilidade individual no sentido amplo ou herdabilidade dos efeitos genotípicos.

Em caso de ausência do efeito de ambiente comum à família, o componente c^2 pode ser ignorado.

1.6. Regressão, Correlação e Coeficiente de Determinação Envolvendo os Valores Genéticos e Fenotípicos

Os coeficientes de regressão, correlação e determinação relativos aos valores genético aditivo e fenotípico são dados por:

$$b_{ay} = \frac{\text{cov} [(\mu + a), y]}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_y^2} = h_a^2: \text{regressão do valor genético aditivo sobre o fenótipo}$$

ou herdabilidade no sentido restrito.

$$r_{ay} = \frac{\text{cov} [(\mu + a), y]}{\sigma_a \sigma_y} = \frac{\sigma_a}{\sigma_y} = h_a: \text{correlação entre o valor genético aditivo e o fenótipo}$$

ou acurácia da avaliação do valor genético aditivo com base no fenótipo.

$$r_{ay}^2 = \frac{\sigma_a^2}{\sigma_y^2}: \text{coeficiente de determinação (quadrado da correlação) do valor genéti-}$$

co aditivo pelo fenótipo ou herdabilidade no sentido restrito.

Os coeficientes de regressão e determinação envolvendo o valor genotípico e o fenotípico são dados por:

$$b_{gy} = r_{gy}^2 = \frac{\sigma_g^2}{\sigma_y^2} = h_g^2 : \text{herdabilidade individual no sentido amplo.}$$

A correlação entre os valores genotípico e fenotípico é dada por: $r_{gy} = h_g$: acurácia da avaliação do valor genotípico com base no fenótipo.

1.7. Parâmetros Genéticos Populacionais

1.7.1. Conceitos

Os parâmetros genéticos populacionais essenciais à aplicação da genética quantitativa no melhoramento são:

Herdabilidade no sentido restrito (h_a^2)

Quadrado da correlação entre o valor genético aditivo e o valor fenotípico, ou regressão do valor genético aditivo sobre o valor fenotípico, ou coeficiente de determinação do valor genético aditivo pelo fenótipo.

Herdabilidade no sentido amplo (h_g^2)

Quadrado da correlação entre o valor genotípico e o valor fenotípico, ou regressão do valor genotípico sobre o valor fenotípico, ou coeficiente de determinação do valor genotípico pelo fenótipo.

Repetibilidade (ρ)

Correlação fenotípica entre medidas repetidas de um mesmo caráter em um mesmo indivíduo, no tempo ou espaço.

Correlação Genética entre Caracteres (r_a)

Medida da magnitude da associação genética entre caracteres em um indivíduo.

Correlação Fenotípica entre Caracteres (r)

Medida da magnitude da associação fenotípica entre caracteres em um indivíduo.

Correlação Fenotípica Permanente (r_{fp})

Medida da magnitude da associação fenotípica entre caracteres em um indivíduo, a qual permanece constante de uma medição a outra, em um indivíduo.

O delineamento de estratégias eficientes de melhoramento depende, fundamentalmente, do conhecimento do controle genético dos caracteres a serem melhorados. Entende-se por controle genético, ou base genética de um caráter,

todos os mecanismos genéticos responsáveis pela herança do caráter, tais como o número de genes que o governam, as ações e efeitos gênicos, herdabilidade, repetibilidade e associações genéticas com outros caracteres.

Dentre estes fatores, os parâmetros genéticos populacionais denominados herdabilidade, repetibilidade e correlação genética entre caracteres são de particular importância para o melhoramento. A herdabilidade diz respeito à proporção relativa das influências genéticas e ambientais na manifestação fenotípica dos caracteres e indica, portanto, o grau de facilidade ou dificuldade para melhorar determinados caracteres. Caracteres com herdabilidade baixa demandarão métodos de seleção mais elaborados do que aqueles com herdabilidade alta.

O parâmetro correlação genética entre caracteres denota o grau de associação genética entre os caracteres, ou seja, quantifica as influências que determinados caracteres exercem sobre outros caracteres. Este parâmetro é muito importante em melhoramento genético, pois significa que, com correlação genética alta, a alteração em um caráter, via seleção, promove alterações significativas em outros caracteres correlacionados a ele.

Outro parâmetro extremamente importante em plantas perenes é o coeficiente de repetibilidade. Este coeficiente mede a capacidade de os organismos repetirem a expressão do caráter, ao longo de vários períodos de tempo, no decorrer de suas vidas. Do ponto de vista prático, suas maiores importâncias são: permitir a determinação do número de medições necessárias para avaliar, com precisão, os valores genéticos aditivos, genotípicos ou fenotípicos permanentes dos indivíduos; possibilitar o uso de suas estimativas nos procedimentos de predição de valores genéticos.

Estimativas de parâmetros genéticos para algumas espécies perenes florrestais, frutíferas e palmáceas são apresentadas no Capítulo 14.

1.7.2. Herdabilidade

O conceito do coeficiente de herdabilidade no sentido restrito foi definido por Lush (1936), em termos de ação aditiva dos genes. De maneira genérica, as definições para a herdabilidade no sentido restrito são:

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_d^2 + \sigma_e^2}$$

Para caracteres avaliados uma vez nos indivíduos e populações experimentais com uma planta por parcela.

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_d^2 + \sigma_{e_p}^2 + \sigma_{e_t}^2}$$

Para caracteres avaliados mais de uma vez nos indivíduos e populações experimentais com uma planta por parcela.

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_d^2 + \sigma_c^2 + \sigma_{e'}^2}$$

Para caracteres avaliados uma vez nos indivíduos e populações experimentais com várias plantas por parcela.

$$h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_d^2 + \sigma_{c'}^2 + \sigma_{e'_p}^2 + \sigma_{e'_t}^2}$$

Para caracteres avaliados mais de uma vez nos indivíduos e populações experimentais com várias plantas por parcela.

As expressões apresentadas referem-se à herdabilidade no estrato denominado bloco. A denotação ' associada aos componentes de variância indica que estes podem se alterar quando são usadas uma planta por parcela (ou uma medição por indivíduo) ou várias plantas por parcela (ou várias medições por indivíduo).

Definições similares se aplicam para a herdabilidade no sentido amplo, bastando substituir σ_a^2 por $\sigma_g^2 = \sigma_a^2 + \sigma_d^2$ no numerador das expressões apresentadas.

1.7.3. Repetibilidade individual

O coeficiente de repetibilidade foi conceituado por Lush (1937) como correlação fenotípica intraclasse e, portanto, no caso de indivíduos refere-se a correlações fenotípicas entre medições repetidas no mesmo indivíduo. Duas medições (y) repetidas no mesmo indivíduo podem ser assim descritas em termos de modelos lineares:

$y = \mu + g + e_p + e_t$, em termos gerais.

$y_1 = \mu + g + e_p + e_{t1}$, para a medição 1.

$y_2 = \mu + g + e_p + e_{t2}$, para a medição 2, em que:

μ : efeito da média geral, fixo.

g : efeito genético, aleatório.

e_p : efeito de ambiente permanente, aleatório.

e_t : efeito de ambiente temporário, aleatório.

A correlação intraclasse ou repetibilidade (ρ) neste caso equivale a:

$$\rho = \frac{COV(y_1, y_2)}{Var(y)} = \frac{COV_g(y_1, y_2) + COV_{ep}(y_1, y_2)}{Var(y)}, \text{ em que:}$$

$$COV_g(y_1, y_2) = COV_g(I, I) = \sigma_i^2 = \sigma_a^2 + \sigma_d^2;$$

$COV_{ep}(y_1, y_2) = \sigma_{ep}^2$: variância ambiental permanente.

$Var(y) = \sigma_y^2$: variância fenotípica.

A repetibilidade é dada por:

$$\rho = \frac{\sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2}{\sigma_y^2}$$

1.7.4. Correlação entre caracteres

Turner & Young (1969) apresentam as seguintes igualdades:

Correlação genética aditiva entre os caracteres x e y

$$r_{a(x,y)} = \frac{COV_{a(x,y)}}{\sigma_{ax}\sigma_{ay}}$$

Correlação fenotípica entre os caracteres x e y

$$r_{(x,y)} = \frac{COV_{a(x,y)} + COV_{e(x,y)}}{\sigma_x\sigma_y} = \frac{COV_{(x,y)}}{\sigma_x\sigma_y}$$

Correlação fenotípica permanente entre os caracteres x e y

$$r_{fp(x,y)} = \frac{COV_{fp(x,y)}}{\sigma_{fp,x}\sigma_{fp,y}}$$

e ainda: $r_{(x,y)} = r_{a(x,y)}h_{ax}h_{ay} + r_{e(x,y)}e_xe_y$

$$r_{e(x,y)} = \frac{COV_{e(x,y)}}{\sigma_{e_x}\sigma_{e_y}}$$

$$e_x^2 = 1 - h_{ax}^2 ; e_y^2 = 1 - h_{ay}^2$$

Esta última expressão para a correlação fenotípica mostra que esta não equivale à soma das correlações genética e ambiental, mas depende, também, das herdabilidades das duas características envolvidas.

Definem-se ainda os seguintes parâmetros:

$cov_{a(x,y)}$: covariância genética aditiva entre os caracteres x e y.

σ_{ax} : desvio padrão genético aditivo para o caráter x.

$cov_{e(x,y)}$: covariância ambiental entre os caracteres x e y.

σ_{ey} : desvio padrão ambiental para o caráter y.

$r_{e(x,y)}$: correlação ambiental entre os caracteres x e y.

$cov_{(x,y)}$: covariância fenotípica entre os caracteres x e y.

σ_y : desvio padrão fenotípico para o caráter y.

$\text{cov}_{fp(x,y)}$: covariância fenotípica permanente entre os caracteres x e y.

σ_{fp_y} : desvio padrão fenotípico permanente para o caráter y.

$\sigma_{fp}^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2$: variância fenotípica permanente.

1.7.5. Parâmetros genéticos e seleção para características de limiar

Para caracteres quantitativos que apresentam distribuição discreta, a análise quantitativa exige considerações adicionais. Características de limiar possuem uma continuidade básica, porém com um limiar que impõe uma descontinuidade sobre a expressão fenotípica. Assim, para estes caracteres, podem ser consideradas duas escalas: a escala base, na qual a variação é contínua e invisível, e a escala visível, na qual a variação é descontínua. Essas duas escalas são conectadas por um limiar ou ponto de descontinuidade (Falconer, 1989).

1.7.5.1. Herdabilidade e seleção

As características de limiar avaliadas pelos escores 0 ou 1, em geral, apresentam distribuição binomial. Considerando esta distribuição, algumas modalidades de análise dessas características poderiam ser utilizadas: (i) análise de variância e estimação de parâmetros genéticos com os dados na escala binomial (valores 0 e 1); (ii) transformação de dados para a escala normal (em geral, via transformação arco-seno da raiz quadrada das proporções observadas) antes da análise e estimativa de parâmetros e (iii) transformação de parâmetros obtidos na escala binomial, para a escala normal por meio da transformação de probitos.

Os parâmetros genéticos obtidos com os dados na escala binomial podem ser utilizados para a predição de ganhos genéticos na própria escala observada. Entretanto, nesta escala existe o inconveniente de se ter dependência entre média e variância (fato que apresenta objeções teóricas, pois viola uma das suposições da análise de variância), de forma que a estimativa da herdabilidade na escala binomial (h_b^2) é amplamente dependente de p, a freqüência ou incidência média do fenótipo desejável na população.

A transformação arco-seno faz o denominador da herdabilidade independente da média p, enquanto a variância genética aditiva mantém-se ainda dependente de p (Robertson & Lerner, 1949). Isto porque, mesmo com a transformação, trabalha-se com os dados fenotípicos na escala visível, e a variância genética aditiva é definida e deve ser medida na escala base (que é mais realística). Na utilização da transformação arco-seno, opera-se ao nível de médias de progénies e, para o cômputo da variância dentro de progénies (dados 0 e 1),

considera-se a variância de amostragem binomial constante de 820,7, conforme Crittenden et al. (1957).

A transformação de probitos (Bliss, 1935; Dempster & Lerner, 1950), por sua vez, realiza diretamente a transformação entre escalas (binomial para a normal e vice-versa) e faz com que a variância genética aditiva seja independente da média p , enquanto o denominador varia de uma maneira perfeitamente conhecida. A variância aditiva na escala binomial equivale a $\sigma_{ab}^2 = z^2 h_n^2$, em que z é a ordenada da curva normal padrão no ponto igual a p e h_n^2 é a herdabilidade aditiva na escala base, escala de probitos ou escala normal padrão. Sendo a variância de uma distribuição Bernoulli (a distribuição Binomial equivale à soma de distribuições Bernoulli) igual $p(1-p)$, tem-se que $h_b^2 = z^2 h_n^2 / [p(1-p)]$ (Dempster & Lerner, 1950). Assim, $h_b^2 = h_b^2 [p(1-p)] / z^2$ fornece a herdabilidade na escala normal, a qual é mais acurada, conforme já relatado no relevante trabalho de Lush et al. (1948). Com a transformação de probitos, verifica-se a igualdade:

$$h_b^2 = \frac{\sigma_{ab}^2}{\sigma_{yb}^2} = \frac{z^2 h_n^2}{p(1-p)} = \frac{h_n^2}{p(1-p)/z^2} = \frac{\sigma_{a\eta}^2}{p(1-p)/z^2}.$$

Esta última igualdade revela que a transformação faz o numerador da herdabilidade independente de p e o denominador variar de uma maneira perfeitamente conhecida dada por $p(1-p)/z^2$. Van Vleck (1972), em estudo de simulação, concluiu que o ajustamento desta teoria é bastante adequado, ou seja, as herdabilidades verdadeiras (escala normal) são apenas ligeiramente superestimadas, quando estimadas por meio de multiplicações da herdabilidade na escala binomial por $p(1-p)/z^2$. Atualmente, as estimativas de h_b^2 são obtidas via metodologia de modelos mistos e, em seguida, transformadas para h_n^2 (Mc Guirk, 1989).

A Tabela 1.1 apresenta valores relativos da fração $z^2 / [p(1-p)]$ em função da variação em p . Estes valores permitem predizer o quanto h_b^2 varia em função da variação em p , para um valor fixo de h_n^2 . A Tabela 1.1 apresenta também valores de $\{z^2 / [p(1-p)]\}(1-p)$ como indicativo da variação nos ganhos genéticos com a seleção individual, em função da variação em p .

Tabela 1.1. Valores relativos da fração $z^2/[p(1-p)]$ e $[z^2/[p(1-p)]](1-p)$ em função da variação em p.

P	z	$z^2/[p(1-p)]$	$[z^2/[p(1-p)]](1-p)$
0,10	0,1758	0,3434	0,31
0,20	0,2803	0,4911	0,39
0,30	0,3485	0,5783	0,40
0,40	0,3867	0,6231	0,37
0,50	0,3989	0,6365	0,32
0,60	0,3867	0,6231	0,25
0,70	0,3485	0,5783	0,17
0,80	0,2803	0,4911	0,10
0,90	0,1758	0,3434	0,03

Verifica-se que h_b^2 é sempre menor que h_n^2 e, h_b^2 assume valores máximos para $p = 0,5$ (h_b^2 equivalerá no máximo a 0,636 da h_n^2) e ao redor desta incidência apresenta valores simétricos. Na escala 0 e 1, o diferencial de seleção é dado por $d_{S_b} = M_S - M_p = 1 - p$, em que M_S e M_p , referem-se à média dos indivíduos selecionados e à média da população, respectivamente. Assim, o ganho genético, computado por $G_S = h_b^2 d_{S_b}$, será maximizado para um mesmo valor de h_b^2 , quando a incidência dos fenótipos desejados for da ordem de 30% (Tabela 1.1) e a proporção selecionada for igual ou inferior a este valor (30%). Se a proporção selecionada for menor que a incidência, o diferencial de seleção na escala binomial continuará o mesmo. Assim, o diferencial de seleção é limitado pela própria incidência. Para características de resistência a parasitas e doenças, isto indica que a seleção será mais favorável quando a taxa de incidência dessas doenças situar-se ao redor de 70% nas populações, fato que pode ser conseguido através de inoculações artificiais (Resende et al., 1993a). Outra alternativa para melhorar a resposta à seleção para características de limiar é o uso da seleção indireta baseada em caracteres com distribuição contínua, correlacionados com essas características.

A Tabela 1.2 apresenta valores de h_b^2 em função da variação em p, para cada valor de h_n^2 .

Tabela 1.2. Valores de h_b^2 em função da variação em p, para cada valor de h_n^2 .

h_n^2	p	h_b^2
0,05	0,05 ou 0,95	0,011
	0,20 ou 0,80	0,025
	0,50	0,032
0,20	0,05 ou 0,95	0,045
	0,20 ou 0,80	0,10
	0,50	0,13
0,50	0,05 ou 0,95	0,11
	0,20 ou 0,80	0,25
	0,50	0,32
0,80	0,05 ou 0,95	0,18
	0,20 ou 0,80	0,40
	0,50	0,50
0,95	0,05 ou 0,95	0,21
	0,20 ou 0,80	0,47
	0,50	0,60

Verifica-se que, quanto mais próximo de 0 ou 1 for p, maior a discrepância entre os valores de h_n^2 e h_b^2 . Constatase, também, que determinada herdabilidade estimada na escala binomial poderá ser aumentada até um certo limite, elevando-se ou diminuindo-se p a 0,5. Por exemplo, para uma herdabilidade verdadeira (h_n^2) de 0,20, uma estimativa de $h_b^2 = 0,10$ obtida com $p = 0,80$ poderá ser elevada a, no máximo, $h_b^2 = 0,13$, fazendo-se $p = 0,50$.

Outra alternativa de análise de variáveis discretas na própria escala binomial foi desenvolvida por Robertson & Lerner (1949), por meio da qual a herdabilidade é estimada por (considerando progêniess de meios-irmãos):

$$h_b^2 = [4 [\chi_{q-1}^2 - (q-1)]]/[n (q-1)], \text{ em que:}$$

χ_{q-1}^2 : qui-quadrado de heterogeneidade, calculado em uma tabela 2 x q;

q: número de progêniess;

n: número médio de indivíduos por progênie.

1.7.5.2. Correlações fenotípicas e genéticas

As correlações fenotípicas entre caracteres avaliados na escala binomial também devem ser convertidas para a escala base ou normal. Para dois caracteres x e y , avaliados na escala binomial, a transformação é dada por Olausson & Ronningen (1975):

$$r_{(x,y)n} = r_{(x,y)b} \left[\frac{p_x(1-p_x)}{z_x^2} \right]^{1/2} \left[\frac{p_y(1-p_y)}{z_y^2} \right]^{1/2}$$

Para o caráter x avaliado na escala binomial e o caráter y avaliado na escala normal, a transformação é dada por:

$$r_{(x,y)n} = r_{(x,y)b} \left[\frac{p_x(1-p_x)}{z_x^2} \right]^{1/2}, \text{ em que:}$$

$r_{(x,y)n}$ e $r_{(x,y)b}$: correlações fenotípicas nas escalas normal e binomial, respectivamente.

p_x e p_y : incidências para os caracteres x e y , respectivamente.

z_x e z_y : ordenadas da curva normal nos pontos p_x e p_y , respectivamente.

As correlações genéticas de dois caracteres avaliados na escala binomial, ou um caráter avaliado na escala binomial e outro na escala normal, são equivalentes nas escalas binomial e normal (Olausson & Ronningen, 1975). Isto porque as correlações genéticas podem ser calculadas por meio da análise de médias de famílias, em que a aproximação da distribuição binomial para a distribuição normal pode ser assegurada.

Deve-se, preferencialmente, trabalhar com os parâmetros na escala normal, devido às vantagens:

- Comparação entre estimativas de parâmetros de diferentes populações e obtidas em diferentes ambientes.
- Maior confiabilidade na predição e explicação da resposta à seleção.
- Obtenção de parâmetros genéticos padrões.
- Obtenção de parâmetros genéticos para formulação de índices de seleção envolvendo simultaneamente variáveis contínuas e discretas, fato que requer transformações das herdabilidades das variáveis discretas e das correlações fenotípicas de ambas.
- Obtenção de parâmetros genéticos para estudos e prática de seleção indireta e seleção usando caracteres auxiliares.
- Predição dos valores genéticos de genitores na escala normal e, posteriormente, transformação dos valores genéticos preditos para a escala binomial.

A abordagem descrita neste tópico pode ser utilizada, também, para características de limiar que apresentam mais de duas classes fenotípicas.

1.7.5.3. Repetibilidade

Para caracteres binários e tendo-se duas avaliações por indivíduo, uma tabela de contingência 2 x 2 sumaria toda a informação amostral. Tal tabela é estruturada da seguinte forma:

Avaliação 1

		0	1	Total
		n ₀₀	n ₁₀	n _{.0}
Avaliação 2	0			
	1	n ₀₁	n ₁₁	n _{.1}
Total		n _{0.}	n _{1.}	n _{..}

Com base nessa tabela de contingência, Lush (1956) apresentou a estatística $\rho_L = n_{11}/n_1 - n_{01}/n_0$ como o estimador para a repetibilidade na escala binomial ou correlação entre avaliações repetidas em um mesmo indivíduo.

Para o caso da variável sobrevivência, podem ser especificadas as quantidades:

n_{11} : número de indivíduos que estavam vivos na primeira avaliação e permaneceram vivos na segunda avaliação, ou seja, é o próprio número de indivíduos vivos na segunda avaliação.

n_1 : soma do número de indivíduos vivos na primeira avaliação e que morreram na segunda avaliação mais o número de indivíduos vivos na segunda avaliação, ou seja, é o próprio número de indivíduos vivos na primeira avaliação.

n_{01} : número de indivíduos que estavam mortos na primeira avaliação e que se tornaram vivos na segunda avaliação, ou seja, zero.

Assim, neste caso, o estimador para a repetibilidade equivale a $\rho_L = N_2/N_1 = p_2/p_1$, em que N_2 e N_1 são os números de indivíduos vivos na segunda e primeira avaliações, respectivamente, e p_2 e p_1 são as porcentagens de sobrevivência na segunda e primeira avaliações, respectivamente.

Considerando ainda a tabela de contingência apresentada, o estimador para a correlação (na escala normal) entre avaliações repetidas apresentada por Plackett (1965) equivale a:

$$\rho_p = \operatorname{sen} \pi/2 \left[\frac{(n_{00} n_{11})^{1/2} - (n_{01} n_{10})^{1/2}}{(n_{00} n_{11})^{1/2} + (n_{01} n_{10})^{1/2}} \right]$$

Considerando a igualdade $\sin \theta = -\cos(\pi/2 + \theta)$, esta correlação é equivalente a:

$$r_p = -\cos [\Psi^{1/2} \pi / (1 + \Psi^{1/2})], \text{ em que:}$$

$$\Psi = n_{00} n_{11} / n_{10} n_{01}.$$

No caso da sobrevivência, n_{01} é igual a zero e, portanto, a expressão r_p não se aplica. Entretanto, para uma variável como resistência a doenças, avaliada em termos de ausência e presença, o estimador r_p pode ser empregado para obtenção da repetibilidade, no caso em que as avaliações 1 e 2 são realizadas sobre o mesmo indivíduo. No caso em que a avaliação 1 é realizada sobre genitores e a avaliação 2 sobre a progénie, o estimador r_p permite obter uma estimativa da herdabilidade do caráter, na escala normal ($r_p = (1/2) h_N^2$).

1.8. Métodos Tradicionais de Estimação de Parâmetros Genéticos

Na situação de dados experimentais balanceados ou aproximadamente平衡ados, os procedimentos tradicionais de estimação paramétrica baseados na técnica de quadrados mínimos (análise de variância e regressão) podem ser aplicados. Vários delineamentos genéticos existem, associados aos diferentes tipos de progêneres, os quais permitem a estimação de parâmetros genéticos. Estes delineamentos foram apresentados detalhadamente por Turner & Young (1969) e Becker (1984). Serão mostrados aqui alguns delineamentos mais comuns.

1.8.1. Análise de regressão

A técnica da análise de regressão permite, em determinadas situações, a estimação de parâmetros genéticos, conforme a Tabela 1.3.

Tabela 1.3. Utilização da análise de regressão na estimação de parâmetros genéticos.

Regressão	Quantidade estimada pelo coeficiente de regressão	Parâmetro	Estimador do parâmetro
Filhos sobre pais	$b = (1/2) \sigma_a^2 / \sigma_y^2$	h_a^2	$\hat{h}_a^2 = 2b$
Filhos sobre a média dos dois pais	$b = \sigma_a^2 / \sigma_y^2$	h_a^2	$\hat{h}_a^2 = b$
Medições adjacentes em um mesmo indivíduo	$b = (\sigma_g^2 + \sigma_{ep}^2) / \sigma_y^2$	ρ	$\hat{\rho} = b$
Média de observações futuras sobre uma observação inicial (para características discretas)	$b = \sigma_{fp}^2 / \sigma_y^2$	ρ	$\hat{\rho} = b$

1.8.2. Análise de variância ou de correlação intraclasse

(a) Herdabilidade no sentido restrito

Um delineamento muito comum no melhoramento de plantas perenes refere-se a observações tomadas em experimentos envolvendo famílias de meios-irmãos, com n indivíduos por família, implantadas no delineamento de blocos ao acaso. Como a covariância entre meios-irmãos fornece a estimativa de uma fração (1/4) da variância aditiva, pode-se estimar a herdabilidade a partir da análise de variância. Considerando a experimentação com uma planta por parcela e ignorando a fonte de variação blocos, tem-se:

Fontes de variação	QM	E (QM)	\hat{h}_a^2
Entre famílias	Q_1	$\sigma^2 + n \sigma_p^2$	$\hat{h}_a^2 = \frac{4 \sigma_p^2}{\sigma^2 + \sigma_p^2} = \frac{4 [(Q_1 - Q_2)/n]}{[Q_1 + Q_2(n-1)]/n} = \frac{4 (Q_1 - Q_2)}{Q_1 + Q_2(n-1)}$
Indivíduos/família	Q_2	σ^2	

Verifica-se também que $h_a^2 = 4 t$, em que $t = (1/4) \sigma_a^2 / \sigma_y^2$ é o coeficiente de correlação fenotípica intraclasse entre meios-irmãos. Os estimadores referentes à experimentação com várias plantas por parcela são apresentados com detalhe no Capítulo 5.

(b) Repetibilidade

Para o cômputo da repetibilidade, a análise de variância pode ser aplicada a m medições realizadas a intervalos regulares de tempo em k indivíduos, conforme o esquema a seguir:

Fontes de variação	QM	E (QM)	$\hat{\rho}$
Medições (M)	Q_1	-	$\hat{\rho} = \frac{(Q_2 - Q_3)/m}{Q_3 + (Q_2 - Q_3)/m} = \frac{Q_2 - Q_3}{Q_2 + Q_3(m-1)}$
Plantas (P)	Q_2	$\sigma^2 + m \sigma_{fp}^2$	
M x P	Q_3	σ^2	

Logicamente, tem-se a igualdade:

$\hat{\rho} = \frac{\hat{\sigma}_{fp}^2}{\hat{\sigma}_y^2} = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2}{\hat{\sigma}_y^2} = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2}{\hat{\sigma}_{fp}^2 + \hat{\sigma}^2}$, em que σ_{fp}^2 é a estimativa da variância permanente entre plantas.

(c) Herdabilidade no sentido amplo

Esquema similar ao descrito em (b) pode ser empregado na estimação da herdabilidade no sentido amplo. Uma análise de variância aplicada a c clones avaliados em b blocos sob uma experimentação com uma planta por parcela, apresenta o seguinte esquema:

Fontes de variação	QM	E (QM)	\hat{h}_g^2
Blocos (B)	Q_1	-	$\hat{h}_g^2 = \frac{(Q_2 - Q_3)/b}{Q_3 + (Q_2 - Q_3)/b} = \frac{Q_2 - Q_3}{Q_2 + Q_3(b-1)}$
Clones (C)	Q_2	$\sigma^2 + b \sigma_g^2$	
B x C	Q_3	σ^2	

Verifica-se também que $h_g^2 = t$, em que $t = \sigma_g^2 / \sigma_y^2$ é o coeficiente de correlação fenotípica intraclasse entre ramos de um mesmo clone.

1.8.3. Estimação da herdabilidade a partir de delineamentos clássicos

Delineamentos de cruzamentos mais elaborados tais como o hierárquico (Delineamento I), o fatorial (Delineamento II) e o dialélico podem ser utilizados para a estimação da herdabilidade (Becker, 1984; Comstock, 1996).

Para o delineamento de cruzamento hierárquico envolvendo f fêmeas por macho e n indivíduos por fêmea, o esquema da análise de variância e estimadores de componentes de variância e herdabilidade, considerando a experimentação com uma planta por parcela e ignorando a fonte de variação blocos, são apresentados a seguir:

Fontes de Variação	QM	E(QM)
Machos	Q_1	$\sigma^2 + n\sigma_{f/m}^2 + nf\sigma_m^2$
Fêmeas/Machos	Q_2	$\sigma^2 + n\sigma_{f/m}^2$
Indivíduos/Fêmeas/Machos	Q_3	σ^2

$$\hat{\sigma}_m^2 = COV_g(MP) = (1/4) \hat{\sigma}_a^2 = (Q_1 - Q_2) / nf .$$

$$\hat{\sigma}_{f/m}^2 = COV_g(IG) - COV_g(MP) = (1/4) \hat{\sigma}_a^2 + (1/4) \hat{\sigma}_d^2 = (Q_2 - Q_3) / n .$$

$\hat{\sigma}^2 = Q_3$: variância residual.

$\hat{h}_a^2 = 4\sigma_m^2 / (\hat{\sigma}_m^2 + \hat{\sigma}_{f/m}^2 + \hat{\sigma}^2)$: herdabilidade no sentido restrito.

$\hat{h}_g^2 = 4\hat{\sigma}_{f/m}^2 / (\hat{\sigma}_m^2 + \hat{\sigma}_{f/m}^2 + \hat{\sigma}^2)$: herdabilidade no sentido amplo.

$COV_g(MP)$: covariância genética entre meios-irmãos paternos.

$COV_g(IG)$: covariância genética entre irmãos germanos.

Para o delineamento de **cruzamento fatorial** envolvendo f fêmeas, m machos e n indivíduos por cruzamento, o esquema da análise de variância e estimadores de componentes de variância e herdabilidade, considerando a experimentação com uma planta por parcela e ignorando a fonte de variação blocos, são apresentados a seguir:

Fontes de variação	QM	E(QM)
Machos (M)	Q_1	$\sigma^2 + n\sigma_{fm}^2 + nf\sigma_m^2$
Fêmeas (F)	Q_2	$\sigma^2 + n\sigma_{fm}^2 + nm\sigma_f^2$
$F \times M$	Q_3	$\sigma^2 + n\sigma_{fm}^2$
Indivíduos/ ($F \times M$)	Q_4	σ^2

$$\hat{\sigma}_m^2 = COV_g(MP) = (1/4) \hat{\sigma}_a^2 = (Q_1 - Q_2) / nf .$$

$$\hat{\sigma}_f^2 = COV_g(MM) = (1/4) \hat{\sigma}_a^2 = (Q_2 - Q_3) / nm .$$

$$\hat{\sigma}_{fm}^2 = COV_g(IG) - COV_g(MP) - COV_g(MM) = (1/4) \hat{\sigma}_d^2 = (Q_3 - Q_4) / n .$$

$\hat{\sigma}^2 = Q_4$: variância residual.

$\hat{h}_a^2 = 4[(\hat{\sigma}_m^2 + \hat{\sigma}_f^2) / 2] / (\hat{\sigma}_m^2 + \hat{\sigma}_f^2 + \hat{\sigma}_{fm}^2 + \hat{\sigma}^2)$: herdabilidade no sentido restrito.

$\hat{h}_g^2 = \hat{h}_a^2 + 4\sigma_{fm}^2 / (\hat{\sigma}_m^2 + \hat{\sigma}_f^2 + \hat{\sigma}_{fm}^2 + \hat{\sigma}^2)$: herdabilidade no sentido amplo.

$COV_g(MM)$: covariância genética entre meios-irmãos maternos.

Para o delineamento de cruzamento dialélico envolvendo g genitores e n indivíduos por cruzamento, o esquema da análise de variância e estimadores de componentes de variância e herdabilidade, considerando a experimentação com uma planta por parcela e ignorando a fonte de variação blocos, são apresentados a seguir:

Fontes de Variação	QM	E(QM)
Capacidade geral de combinação	Q_1	$\sigma^2 + n\sigma_{sca}^2 + n(g-2) \sigma_{gca}^2$
Capacidade específica de combinação	Q_2	$\sigma^2 + n\sigma_{sca}^2$
Indivíduos/cruzamento	Q_3	σ^2

$$\hat{\sigma}_{gca}^2 = COV_g(MP) = COV_g(MM) = (1/4) \hat{\sigma}_a^2 = (Q_1 - Q_2)/[n(g-2)];$$

$$\hat{\sigma}_{sca}^2 = COV_g(IG) - COV_g(MP) - COV_g(MM) = (1/4) \hat{\sigma}_d^2 = (Q_2 - Q_3)/n;$$

$\hat{\sigma}^2 = Q_3$: variância residual.

$\hat{h}_a^2 = 4\sigma_{gca}^2 / (\hat{\sigma}_{gca}^2 + \hat{\sigma}_{sca}^2 + \hat{\sigma}^2)$: herdabilidade no sentido restrito.

$\hat{h}_g^2 = \hat{h}_a^2 + 4\sigma_{sca}^2 / (\hat{\sigma}_{gca}^2 + \hat{\sigma}_{sca}^2 + \hat{\sigma}^2)$: herdabilidade no sentido amplo.

Os modernos métodos de estimação de componentes de variância pela metodologia de modelos mistos (por máxima verossimilhança restrita ou análise bayesiana) são apresentados em capítulos específicos.

1.8.4. Estimação do coeficiente de variação genética

O coeficiente de variação genética é um parâmetro relevante nos estudos de genética quantitativa uma vez que permite inferir sobre a magnitude da variabilidade presente nas populações e em diferentes caracteres. Sendo uma estatística expressa em relação à média das populações e dos caracteres, suas estimativas permitem comparar níveis de variabilidade presentes em diferentes populações, ambientes e caracteres. Os estimadores do referido coeficiente de variação são dados por:

$\hat{cv}_a = \hat{\sigma}_a / \bar{x}$: estimador do coeficiente de variação genética aditiva.

$\hat{cv}_g = \hat{\sigma}_g / \bar{x}$: estimador do coeficiente de variação genotípica.

\bar{x} : estimativa da média populacional para o caráter em questão.

1.9. Interação Genótipo x Ambiente

1.9.1. Conceitos

Na análise de um grupo de indivíduos avaliados em um ambiente ou local, o modelo para o valor genotípico é:

$$y = \mu + g + e$$

Quando grupos de indivíduos são avaliados em mais de um local, o modelo para o valor fenotípico deve ser expandido para:

$y = \mu^* + \ell + g^* + g^*\ell + e$, em que:

μ^* : efeito da média geral livre do efeito de local.

g^* : efeito genotípico, livre da interação genótipo x ambiente.

ℓ : é o efeito de local ou macroambiente.

$g^*\ell$: é o efeito da interação genótipo x ambiente.

Comparando este modelo com o anterior, têm-se as igualdades: $\mu = \mu^* + \ell$; $g = g^* + g^*\ell$. Assim, verifica-se que quando a avaliação é realizada em um único ambiente, a média genotípica é inflacionada pelo efeito de locais e o efeito genotípico, pelo efeito da interação genótipo x ambiente. O efeito da interação genótipo x ambiente é decorrente do comportamento diferencial dos diferentes genótipos nos diferentes ambientes e pode indicar que os melhores indivíduos em um ambiente podem não sê-lo em outro ambiente. Assim, este pode ser um complicador na seleção, se não for considerado adequadamente.

Nesta situação, a variação fenotípica é dada por:

$$\sigma_y^2 = \sigma_{g^*}^2 + \sigma_{g^*\ell}^2 + \sigma_e^2 \text{ (considerando fixo o efeito de locais).}$$

No caso de dois ambientes, as herdabilidades relativas às várias modalidades de seleção são dadas conforme a Tabela 1.4, obtida com base em Resende (1989).

Tabela 1.4. Herdabilidades associadas à seleção em dois ambientes.

Ambiente de seleção	Ambiente onde se deseja o ganho	Herdabilidade*
x	x	$h_{x/x}^2 = \frac{\sigma_{g_x}^2}{\sigma_x^2}$
x	y	$h_{y/x}^2 = \frac{\sigma_{g^*}^2}{\sigma_x^2}$
x	Média dos ambientes x e y	$h_{(x,y)/x}^2 = \frac{(1/2)(\sigma_{g^*}^2 + \sigma_{g_x}^2)}{\sigma_x^2}$
Média dos ambientes x e y	x	$h_{x/(x,y)}^2 = \frac{(1/2)(\sigma_{g^*}^2 + \sigma_{g_x}^2)}{\sigma_{(x,y)}^2}$
Média dos ambientes x e y	Média dos ambientes x e y	$h_{((x,y)/(x,y))}^2 = \frac{(1/4)(\sigma_{gx}^2 + \sigma_{gy}^2 + 2\sigma_{g^*}^2)}{\sigma_{(x,y)}^2} = \frac{\sigma_{g^*}^2 + (1/2)\sigma_{g^*\ell}^2}{\sigma_{(x,y)}^2}$

* σ_x^2 e $\sigma_{(x,y)}^2$: variâncias fenotípicas no ambiente x e na média dos ambientes x e y, respectivamente.

σ_{gx}^2 : variância genética no local x.

$\sigma_{g^*\ell}^2$: variância da interação genótipos com ambientes.

$$\sigma_{g^*}^2 + \sigma_{g^*\ell}^2 = (1/2)(\sigma_{gx}^2 + \sigma_{gy}^2).$$

Verifica-se que quando um material é avaliado em um ambiente e utilizado no mesmo ambiente, o efeito da interação é capitalizado na seleção, ou seja, o valor genotípico dos indivíduos selecionados e utilizados no local ℓ é dado por: $\mu^* + \ell + g^* + g^{*\ell}$.

Quando o material é avaliado em um ambiente e utilizado em outro, o valor genotípico no outro ambiente é menor, eqüivalendo $\mu^* + g^*$. Assim, a interação pode ser favorável ou desfavorável à seleção, dependendo da estratégia de utilização do material selecionado.

1.9.2. Correlação genética através dos ambientes e número de locais de experimentação

O componente de variância da interação genótipo x ambiente pode ser assim desdobrado (Robertson, 1959):

$$\sigma_{g^{*\ell}}^2 = \frac{1}{2}(\sigma_{gx} - \sigma_{gy})^2 + (1 - r_{gxy}) \sigma_{gx} \sigma_{gy}, \text{ em que:}$$

$\frac{1}{2}(\sigma_{gx} - \sigma_{gy})^2$: parte simples da interação, explicada pela mudança de variação genética (heterogeneidade de variâncias) dos materiais nos diferentes ambientes.

$(1 - r_{gxy}) \sigma_{gx} \sigma_{gy}$: parte complexa da interação, advinda da falta de correlação genética entre o desempenho do material genético de um ambiente para outro. É a parte problemática da interação, significando que o material bom em um ambiente pode não sê-lo em outro, quando r_{gxy} é baixa.

r_{gxy} : correlação genética entre o desempenho do material genético de um ambiente para outro.

σ_{gx} : desvio padrão genético aditivo no ambiente x.

Pela expressão de $\sigma_{g^{*\ell}}^2$ verifica-se que, quando a correlação genética através dos ambientes equivale a 1, a variância da interação genótipo x ambiente refere-se, exclusivamente, à heterogeneidade de variâncias.

A correlação r_{gxy} pode ser estimada pelas expressões:

$$\hat{r}_{gxy} = \frac{\sigma_{g^*}^2}{\sigma_{gx} \sigma_{gy}} \text{ ou}$$

$$\hat{r}_{gxy} = \frac{\sigma_{g^*}^2}{\sigma_{g^*}^2 + \sigma_{g^{*\ell}}^2 - 0,5 (\sigma_{gx} - \sigma_{gy})^2} \text{ ou}$$

$\hat{r}_{gxy} = r_{xy} / (h_x h_y)$, em que r_{xy} é a correlação fenotípica entre o desempenho do material genético de um ambiente para o outro e h_x e h_y são as raízes quadradas das herdabilidades nos ambientes x e y, respectivamente. Esta expressão é válida, pois a covariância fenotípica (numerador da correlação fenotípica) neste caso é apenas de natureza genética.

É importante relatar que r_{gxgy} (também denominada correlação genética do tipo B) pode ser menor que zero, quando o seu numerador (que é uma covariância) for menor que zero.

Para situações que envolvem mais que dois locais, a correlação genética pode ser estimada por:

$$\hat{r}_{gg} = \frac{\sigma_{g^*}^2}{\sigma_{g^*}^2 + \sigma_{g^*\ell}^2 - V(\sigma_{gi})}, \text{ em que:}$$

$V(\sigma_{gi})$: variância da escala genética ou variância dos desvios padrões genéticos nos ambientes. Em caso de variações genéticas de magnitudes similares nos ambientes, este valor torna-se nulo.

Na estimação de r_{gg} e r_{gxy} , o conceito do coeficiente de correlação genética intraclasse pode ser também utilizado (Dickerson, 1962). Considerando homogeneidade de variância genética ($V(\sigma_{gi}) = 0$) nos ambientes, tem-se que:

$$\hat{r}_{gxy} = \frac{\sigma_{g^*}^2}{\sigma_{gx} \sigma_{gy}} = \frac{\sigma_{g^*}^2}{\sigma_{g^*}^2 + \sigma_{g^*\ell}^2}$$

Este estimador subestima a correlação genética se houver heterogeneidade de variâncias e, neste caso, é um estimador conservador do ponto de vista do melhorista, já que superestimar (diminuindo a estimativa da correlação genética) o efeito da interação e, consequentemente, levará o melhorista a se prever contra ela. Assim, r_{gg} (assumindo $V(\sigma_{gi}) = 0$) é o limite inferior da correlação genética esperada para um caráter avaliado em parentes em dois diferentes ambientes. Equivale também a um estimador exato de r_{gxy} quando se corrige (padronização), previamente, a heterogeneidade de variância dos dados. Este estimador pode ser aplicado facilmente a partir de uma análise de variância conjunta, segundo um modelo aleatório, como no esquema mostrado a seguir.

Fontes de variação	E(QM)	Teste F
Ambientes (L)	-	-
Genótipos (G)	$\sigma^2 + n \sigma_{g^*\ell}^2 + n\ell \sigma_{g^*}^2$	F
G x L	$\sigma^2 + n \sigma_{g^*\ell}^2$	F*
Resíduo	σ^2	-

n e ℓ : número médio de indivíduos por progênies (genótipos) e número de ambientes de avaliação, respectivamente.

A partir desta análise, têm-se os seguintes estimadores (Resende, 1998):

$$\hat{\sigma}_{g^*\ell}^2 / \hat{\sigma}_{g^*}^2 = \frac{(1 - 1/F^*)}{(F - 1)} \ell = \frac{(F^* - 1)}{(F - 1)} \frac{\ell}{F^*}$$

$$\hat{r}_{gg} = \frac{\sigma_{g^*}^2}{\sigma_{g^*}^2 + \sigma_{g^*\ell}^2} = \frac{(F - 1)}{(F - 1) + (\ell - \ell/F^*)} = \frac{1}{1 + \frac{(F^* - 1)}{(F - 1)} \frac{\ell}{F^*}}$$

As correlações r_{gg} e r_{gxgy} podem ser expressas alternativamente em função da proporção $P = \sigma_{g^*\ell}^2 / \sigma_{g^*}^2$, por meio de $r_{gg} = \sigma_{g^*}^2 / (\sigma_{g^*}^2 + P \sigma_{g^*}^2) = 1/(1+P)$. Com $P = 0,5$, tem-se $r_{gg} = 0,67$, que é um valor alto de correlação genética. Dessa forma, pode-se inferir que quando a proporção variância da interação/variância genética livre da interação for inferior a 0,5, a interação não é problemática para o melhorista, pois conduzirá a uma alta correlação. Quando $P > 0,5$, a interação poderá ser problemática para o melhorista, implicando perdas de ganho com a seleção indireta (seleção em um local visando a ganho em outro). Tem-se também a igualdade $P = \sigma_{g^*\ell}^2 / \sigma_{g^*}^2 = \frac{1 - r_{gg}}{r_{gg}}$.

A eficiência (E_f) da seleção baseada na média de vários (ℓ) ambientes em relação à seleção em um só ambiente pode ser inferida (para herdabilidades ao nível de médias, similares e tendendo a 1 nos vários ambientes) pela expressão: $E_f = [\ell / [1 + (\ell - 1) r_{gg}]]^{1/2}$. Esta expressão é função de $\left[\frac{h_{(x,y)/(x,y)}^2}{h_{(x,y)/x}^2} \right]^{1/2}$, em que estas herdabilidades estão definidas na Tabela 1.4. A expressão de E_f mede a eficiência do uso de ℓ locais ao invés de um, visando a ganho na média dos ℓ locais.

Com base nesta expressão, foi estabelecida a Tabela 1.5.

Tabela 1.5. Eficiência (em termos de ganho genético na média dos locais) da utilização de ℓ locais de avaliação dos materiais genéticos em vez de apenas um, para diferentes valores da correlação genética (r_{gg}) envolvendo a performance do germoplasma nos ambientes.

r_{gg}	ℓ	E
0,95	2	1,01
	3	1,02
0,90	2	1,03
	3	1,04
0,85	2	1,04
	3	1,05

Continua...

Tabela 1.5. Continuação

r_{gg}	ℓ	E
0,80	2	1,05
	3	1,07
0,75	2	1,07
	3	1,10
0,70	2	1,08
	3	1,12
0,65	2	1,10
	3	1,14
0,60	2	1,12
	3	1,17
0,55	2	1,14
	3	1,20
0,50	2	1,15
	3	1,22
0,40	2	1,20
	3	1,29
0,30	4	1,35
	2	1,24
0,30	3	1,37
	4	1,45
	5	1,51

Verifica-se pela Tabela 1.5 que, quando a correlação genética for igual ou superior a 0,70, o ganho em eficiência, pela utilização de mais de um local de experimentação, é inferior a 10%. Se a correlação genética for superior a 0,80, o ganho em eficiência é inferior a 5%. Por outro lado, a utilização de três em vez de dois locais parece ser recomendada somente quando a correlação (estimada com base em três ou mais locais) for inferior a 0,5. Maiores detalhes sobre a interação genótipo x ambiente e eficiência seletiva são apresentados no capítulo 19, tópicos 19.10 e 19.11.

1.9.3. Estabilidade dos valores genotípicos preditos

A capacidade de os materiais genéticos se comportarem bem em uma grande amplitude de condições ambientais pode ser fator essencial em um programa de melhoramento genético. Neste contexto, o estudo da estabilidade dos

materiais genéticos torna-se relevante. De maneira geral, um material é considerado estável quando apresenta pequenas variações no seu comportamento geral quando é avaliado em diversas condições de ambiente. Existem muitos métodos de avaliação da estabilidade fenotípica (Vencovsky & Barriga, 1992; Ramalho et al., 1993; Cruz & Regazzi, 1994).

Será abordada nesta publicação a estabilidade genotípica ou genética, ou seja, a estabilidade dos valores genéticos preditos para um mesmo material genético (genitor, clone, família, indivíduos) em diferentes ambientes. Para tanto, será descrito o método de Wricke (Wricke & Weber, 1986).

O parâmetro de estabilidade associado ao método de Wricke é denominado ecovalência (W_i) e equivale à decomposição da soma de quadrados da interação genótipos x ambientes (da análise de variância conjunta) em partes deviadas a genótipos isolados. O estimador de W_i é dado por:

$$\hat{W}_i = \sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{...})^2, \text{ em que:}$$

Y_{ij} : valor genético predito para o material i no ambiente j.

$\bar{Y}_{i.}$: valor genético predito, médio para o indivíduo i nos ambientes.

$\bar{Y}_{.j}$: valor genético predito, médio para o ambiente j.

$\bar{Y}_{...}$: média geral dos valores genéticos preditos em todos os ambientes.

Dessa forma, a soma de quadrados da interação genótipo x ambiente é dada por $SQ_{GxL} = \sum_i W_i$ e a contribuição percentual de cada genótipo para a interação, por $(W_i / \sum_i W_i) \times 100$.

Um teste estatístico para \hat{W}_i , baseado na distribuição F de Snedecor foi apresentado por Lisbão Junior (1991). Tal teste compara a contribuição W_i de cada genótipo para a interação com o remanescente da interação. Assim:

$SQW_i = \frac{g W_i}{g-i}$: soma de quadrados da ecovalência do genótipo i, com

$$(SQW_i / \sigma^2) \sim \chi^2(\ell-1).$$

$SQR = SQ_{GxL} - SQW_i$: remanescente da soma de quadrados da interação, com $(SQR / \sigma^2) \sim \chi^2[(g-2)(\ell-1)]$.

$QMW_i = \left(\frac{g W_i}{g-i} \right) / (\ell-1)$: quadrado médio da ecovalência do genótipo i.

$QMR = \left(\sum_i W_i - W_i \frac{g}{g-1} \right) / [(g-2)(\ell-1)]$: quadrado médio da interação remanescente.

$$F_i = \frac{QMW_i}{QMR} = \frac{(g-2) g W_i}{(g-1) \sum_i W_i - g W_i} \sim F [(\ell-1), (g-2)(\ell-2)], \text{ em que:}$$

g e ℓ : número de genótipos e locais avaliados, respectivamente.

σ^2 : variância média do erro de predição dos valores genéticos. Este componente de variância não é necessário, pois é cancelado quando se faz a razão entre as distribuições qui-quadrado para obtenção da distribuição F.

A metodologia apresentada permite uma descrição contínua (por meio das contribuições percentuais) da estabilidade genética dos materiais, fornecendo uma distribuição coerente com a natureza quantitativa do caráter estabilidade. Adicionalmente, permite uma classificação discreta dos genótipos em estáveis e não estáveis, empregando-se para isto uma distribuição F com $(\ell-1)$ e $(g-2)$ ($\ell-1$) graus de liberdade.

Outras estatísticas relevantes que podem ser obtidas em função da ecovariância são o desvio padrão do valor genético médio prevido (\bar{Y}_i) dado por $s(\bar{Y}_i) = [W_i / (\ell-1)]^{1/2}$ e o intervalo de confiança do referido valor genético, dado por $\bar{Y}_i \pm t s(\bar{Y}_i)$ em que t é um valor tabelado associado à distribuição t de Student ($t = 1,96$ para um intervalo com 95% de confiança).

1.10. Tamanho Efetivo Populacional e Endogamia

1.10.1 Conceito fundamental

O conceito de tamanho efetivo populacional (N_e) foi introduzido por Sewall Wright em 1931. Este autor denomina “população ideal” uma população composta por N indivíduos diplóides, reconstituída a cada geração por uma amostra aleatória de $2N$ gametas. Definiu-se, então, o N_e como sendo o tamanho da população ideal que apresenta a mesma taxa de aumento na homozigose ou deriva genética, apresentada pela população real em consideração. Uma definição minuciosa do N_e é dada por Freire-Maia (1974): “corresponde a uma população ideal de tamanho invariável, com número igual de machos e fêmeas, destituída de endocruzamento em taxa apreciável e em que cada indivíduo tem a mesma probabilidade de deixar progênie de tamanho igual, de tal forma que o número de descendentes por genitor, se distribui segundo a série de Poisson”. Esta última consideração implica $N_e = N = N_r$, em que N_r é a população reprodutora.

Em termos genéricos, pode-se dizer que o tamanho efetivo populacional refere-se ao tamanho genético de uma população reprodutiva e não ao número de indivíduos que a compõe. Assim, diz respeito à representatividade genética de amostras de animais, plantas e sementes. Conforme implícito no próprio conceito de Wright, na situação de um tamanho efetivo pequeno, um reduzido número de indivíduos participa efetivamente do intercruzamento com vistas à regeneração da nova população, conduzindo a ocorrência de dois eventos: mudança aleatória das freqüências alélicas (oscilação ou deriva genética); aumento da endogamia na nova população.

A deriva genética pode ser definida como o processo dispersivo que surge em pequenas populações em consequência da amostragem. Considere-se uma

população com freqüência p_o para um alelo neutral a seleção. Tomando-se várias amostras aleatórias desta população, a freqüência desse alelo, em cada amostra, distribui-se aleatoriamente, ao redor de \bar{p}_1 , a freqüência alélica média nas amostras.

Tem-se que:

$$E(\bar{p}_1) = p_o$$

$\sigma_{\bar{p}_1}^2 = \frac{p_o(1-p_o)}{2N}$: variância de uma proporção, sendo $2N$ o tamanho da amostra de alelos da população base diplóide, ou seja, $2N_e$. Corresponde à variância de amostragem da distribuição binomial.

Assim, embora $\bar{p}_1 = p_o$, uma particular amostra poderá apresentar freqüência alélica diferente de p_o . Pela expressão da variância das freqüências alélicas, verifica-se que quanto menor o tamanho da amostra (menor N_e), maiores serão os desvios em relação à freqüência original, e uma amostra representará pobramente a população.

A deriva genética causa mudanças nas freqüências alélicas e essas mudanças podem ser preditas apenas em termos de quantificação, pela variância das freqüências alélicas, mas não em termos de direção. Isto porque as alterações que ocorrem nas freqüências alélicas são totalmente aleatórias, exatamente por causa da amostragem de gametas para a formação da nova geração. A deriva conduz, então, à fixação e a perdas de alelos, a uma taxa de $1/(4N_e)$ fixações e $1/(4N_e)$ perdas por geração, ou seja, a uma redução da variabilidade em $1/(2N_e)$.

A partir da amostragem inicial, amostragens adicionais em cada subpopulação contribuem para aumentar a diferenciação entre as subpopulações. Isto ocorre em função da dispersão adicional sobre a freqüência alélica p_1 , particular a cada subpopulação e, portanto, diferente de p_o e \bar{p}_1 .

Assim, em cada geração n de amostragem, a variância da freqüência alélica entre as subpopulações equivale (Falconer, 1989):

$$\sigma_{pn}^2 = p_o(1-p_o)[1 - (1 - 1/(2N))^n]$$

1.10.2. Diferentes conceitos do tamanho efetivo

Crow & Denniston (1988) relatam a existência de vários tipos de N_e , os quais conduzem, às vezes, a expressões de cálculo diferenciadas. Basicamente, três tipos são descritos:

1. **Tamanho efetivo de endogamia:** é o próprio conceito de Wright, baseado na probabilidade de homozigose devido à ancestralidade comum. Portanto, fornece uma medida da taxa de perda de indivíduos heterozigotos na população.
2. **Tamanho efetivo de variância:** baseado na deriva alélica, ou seja, na variância das freqüências alélicas por geração de amostragem. Portanto, fornece uma medida da taxa de variação genética na população.

- 3. Tamanho efetivo de extinção (ou de autovalores):** baseado na taxa assintótica de decréscimo de locos segregantes (definido, inicialmente, por Ewens, 1979).

O tamanho efetivo de variância é mais útil em conservação genética, pois está relacionado com a conservação da variabilidade genética. O tamanho efetivo de endogamia é mais apropriado no cômputo de coeficientes de endogamia, apresentando também utilidade à conservação genética. Por sua vez, o tamanho efetivo de extinção é menos usado, podendo, entretanto, ser maximizado em estratégias de conservação delineadas para resgatar espécies, por meio da maximização do número de eqüivalentes fundadores (Templeton & Read, 1994). Neste documento são considerados apenas o tamanho efetivo de variância e de endogamia.

1.10.3. Conceitos aplicados do tamanho efetivo

Em termos aplicados, o tamanho efeito populacional pode ser calculado de duas maneiras: em relação à geração imediatamente anterior, sendo denominado N_e ou em relação a uma geração inicial de referência, sendo denominado N'_e .

O N_e pode assumir quaisquer valores, tendo o ∞ como limite superior. N'_e assume valores cada vez menores com o decorrer das gerações, tendo como limite superior o próprio N_e .

O N_e é particularmente importante para populações de melhoramento. Isto porque se trabalha com a probabilidade de fixação dos alelos favoráveis e, portanto, o N_e deve ser mantido em um nível seguro a cada passagem de geração, visando evitar a perda desses alelos em cada passo de amostragem.

N'_e deve ser empregado em populações de conservação de germoplasmas, pois, nesse caso, objetiva-se manter todo ou pelo menos grande parte do conjunto gênico original.

Neste contexto, é importante relatar que a redução em N'_e , ocorre mesmo que seja mantido constante o tamanho da amostra ao longo de gerações sem seleção. Isto pode ser constatado pela própria expressão de cálculo do N'_e , a qual equivale à média harmônica dos N_e observados em cada passo amostral n (Vencovský, 1988):

$$\frac{1}{N'_e} = \left[\frac{1}{N_{e_1}} + \dots + \frac{1}{N_{e_n}} \right] \frac{1}{n}$$

Neste caso, N'_e depende mais dos valores mais baixos de N_{ei} .

Outro conceito aplicado de tamanho efetivo refere-se àquele aplicável a amostras compostas por subpopulações amostradas, independentemente, em uma população base. Conforme relatado por Vencovský (1988), reunindo R amostras independentes e iguais em quantidades de propágulos ou indivíduos, cada uma com tamanhos efetivos arbitrários $N_{e1}, N_{e2}, \dots, N_{eR}$, o tamanho efetivo da amostra composta (N_{et}) é dado por:

$$N_{et} = \frac{R^2}{\sum_j^R \frac{1}{N_{ej}}} = R \cdot \bar{N}_e, \text{ em que:}$$

N_{ej} : tamanho efetivo de cada amostra.

\bar{N}_e : média harmônica de N_{ej} .

$j = 1, 2, \dots, R$.

Outra maneira de maximizar N_{et} , apresentada por Resende & Vencovsky (1990), é a tomada de quantidades de propágulos ou indivíduos de cada amostra, proporcionais ao respectivo tamanho efetivo delas. Neste caso, $N_{et} = \sum_j N_{ej} \cdot$

O cômputo de N'_e e N_{et} , para quaisquer situações e sistemas reprodutivos, é dependente apenas do conhecimento do N_e , cujas expressões de cálculo serão particularizadas em tópico seguinte.

1.10.4. Conceito matemático do tamanho efetivo

A teoria matemática do N_e postula que uma população é composta por N indivíduos, os quais contribuem com variáveis números (k) de gametas para a geração seguinte, dentre um total de $2N$. Nesta situação têm-se:

- Número médio de gametas contribuído por indivíduo: $\bar{k} = 2$
- Variância do número de gametas contribuídos pelos N indivíduos:

$$\sigma_k^2 = \left[\sum_{i=1}^N (k_i - \bar{k})^2 \right] / N$$

- Expressão clássica do tamanho efetivo: $N_e = \frac{4N-2}{\sigma_k^2 + 2}$

Observa-se pela expressão do N_e , que o fator que, em geral, o torna menor que o número absoluto de indivíduos da população é exatamente σ_k^2 , ou seja, a variância do número de gametas contribuídos pelos genitores. Com $\sigma_k^2 = 0$ (igual tamanho de família) e dois indivíduos sobreviventes por genitor, $N_e = 2N-1$, ou seja, o tamanho efetivo praticamente duplica em relação ao original. Percebe-se, então, que a contribuição eqüitativa de indivíduos por genitor é sempre desejável, quando se procura maximizar o N_e .

Considerando que os genitores têm iguais probabilidades de contribuir para a próxima geração (distribuição de Poisson), $\sigma_k^2 = \bar{k} = 2$ e $N_e = N + \frac{1}{2}$, ou seja, praticamente como na população idealizada. Supondo tamanhos constantes das populações reprodutoras ao longo das gerações, implica $k = 2$ e a expressão clássica pode ser usada de forma generalizada, de forma que N eqüivalerá ao número de genitores selecionados, para espécies monóicas e será $N = N_f + N_m = 2 N_f$ para espécies dióicas, quando $N_f = N_m$ e σ_k^2 é o mesmo para

os dois sexos. N_f e N_m referem-se ao número de genitores femininos e masculinos, respectivamente.

A expressão clássica pode ser rearranjada em termos mais genéricos (para qualquer \bar{k}) e em função de amostras (correção gaussiana) e não da população total, como:

$$N_e = \frac{N\bar{k}(N\bar{k} - 1)}{(N-1)\sigma_k^2 + N\bar{k}(\bar{k} - 1)}$$

Desprezando-se a correção para tamanho de amostra finito, ou seja, considerando $[(N - 1)/N] = 1$, tem-se:

$$N_e = (N\bar{k} - 1) / (\sigma_k^2 / \bar{k} + \bar{k} - 1) = (N\bar{k}^2 - \bar{k}) / (\sigma_k^2 + \bar{k}^2 - \bar{k}).$$

Esta é a expressão geral do N_e para espécies monóicas, a qual pode ser utilizada para derivar as expressões adequadas às espécies dióicas.

1.10.5. Tamanho efetivo com sobreposição de gerações

Para uma população com tamanho e proporção de sexo constantes e com estável distribuição de idade, o tamanho efetivo com sobreposição de gerações equivale àquele para gerações discretas com tamanho populacional constante e mesma variância do número de indivíduos coetâneos (Hill, 1979).

Uma vez que os indivíduos diferem em tempo de vida, alguns têm a chance de se reproduzir mais vezes que outros, conduzindo ao aumento da variância do tamanho de família e, consequentemente, reduzindo o N_e . O problema de estimação do N_e nesta situação refere-se à determinação de N na expressão clássica do N_e . Desde que a população tenha uma estável estrutura de idade, N pode ser encontrado conforme descrito a seguir:

- Deve-se conhecer o número (N_c) de indivíduos nascidos em dado intervalo de tempo (coetâneos), o qual é dado por $N_c = N_T/E$, onde N_T é o número sobrevivente a qualquer tempo e E é a expectativa de vida ou idade de morte.
- Deve-se conhecer, também, o intervalo de geração L, que equivale à idade média de reprodução.
- Então: $N = N_c L$ e pela expressão clássica com $\bar{k} = 2$ tem-se $N_e = 4N_c L / (\sigma_k^2 + 2)$, aproximadamente, em que σ_k^2 é a variância do tamanho de família.

1.10.6. Endogamia devida ao pequeno tamanho efetivo

O coeficiente de endogamia foi definido inicialmente por Wright como correlação entre gametas que se unem, e Malecot o definiu como a probabilidade de que os dois alelos de um loco em um indivíduo sejam idênticos por descendência. Logo, está implícito que o termo endogamia refere-se ao processo de cruzamento entre indivíduos mais parentados entre si do que o são, membros

aleatórios da população (Kempthorne, 1957) e que tem por consequência um aumento da homozigose da população. Dessa forma, é um conceito relativo à população ancestral.

Considere-se uma população de referência sendo diplóide, monóica (com autofecundação permitida) e não endógama. Todos os genes em um loco são não idênticos por descendência, de forma que numa população de N indivíduos, cada loco produz, aleatoriamente, $2N$ diferentes tipos de gametas igualmente, com frequência $1/(2N)$, para produzir os indivíduos da próxima geração.

Percebe-se, então, que a probabilidade de um indivíduo ser fecundado, a partir de um gameta oriundo dele mesmo é $1/(2N)$. Assim, quanto menor for o tamanho N da população reprodutiva, maior será a endogamia na geração subsequente.

O coeficiente de endogamia devido ao pequeno tamanho populacional, quantificado em termos de aumento da homozigose ou perda de heterozigose, equivale a $F = 1/(2N)$, em que N refere-se a indivíduos não aparentados na população ideal de referência, ou seja, N é igual ao número efetivo de indivíduos que participam da reprodução. Assim, de maneira generalizada, é melhor definir $F = 1/(2N_e)$, onde N_e é o tamanho efetivo populacional. Outra maneira de obter $F = 1/(2N)$ é baseada na probabilidade de dois gametas idênticos se unirem, ou seja, $F = 2N(1/(2N))^2 = 1/(2N)$.

A população reprodutiva da geração 1 (oriunda da população original ou geração 0) novamente é finita e nova endogamia será acrescida à prévia, na 2ª geração. Esta endogamia nova ocorre também na proporção $1/(2N_e)$, e a proporção restante ($1-1/(2N_e)$) em parte é endógama devido à endogamia F_1 precedente, ou seja, devido a alelos idênticos a partir da origem 0 e não 1. Assim, a endogamia da geração 2 será $F_2 = \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right)F_1$, ou seja, a soma das probabilidades dos dois eventos mutuamente exclusivos.

Por extensão, pode-se obter o coeficiente de endogamia em uma geração qualquer n , pela expressão: $F_n = \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right)F_{n-1}$, ou, de maneira mais geral $F_n = \Delta F + (1 - \Delta F)F_{n-1}$, em que $\Delta F = \frac{1}{2N_e} = \frac{F_n - F_{n-1}}{1 - F_{n-1}}$ é o acréscimo na endogamia a partir de uma geração de cruzamentos aleatórios em uma população finita.

Por esta expressão, verifica-se que se não existe endogamia em determinada geração, ou seja, se o tamanho populacional for aumentado substancialmente, não existe ΔF , mas toda a endogamia (F_{n-1}), obtida, previamente, permanece.

1.10.7. Efeitos da endogamia na média (depressão endogâmica) e variâncias populacionais

Com endogamia, a média genotípica populacional é reduzida desde que haja variabilidade genética e algum nível de dominância alélica. O fenômeno associado à redução na média é denominado depressão endogâmica, o qual resulta da mudança das freqüências genotípicas da população, após esta ter sofrido endogamia.

O termo depressão endogâmica caracteriza-se pelos efeitos da perda de heterozigose, comparando-se a população não endógama (com $F = 0$) com a população totalmente endogâmica (com $F = 1$). Assim, com uma geração de autofecundação em uma população não endógama tem-se $F = 1/2$ e, consequentemente, metade da depressão endogâmica.

A homozigose produzida pela endogamia pode ser útil na detecção e eliminação de alelos recessivos deletérios e semideletérios. Assim, métodos de melhoramento que levem à ocorrência de endogamia podem ser interessantes como forma de reduzir a carga genética existente em populações panmíticas.

Importante faz-se, ainda, relatar que a endogamia não altera as freqüências alélicas, de forma que por si só não fornece ao melhorista o que ele deseja. Sua utilização deve ser então acompanhada de seleção, que é a única força capaz de conduzir ao melhoramento.

Considerando toda uma população, a endogamia aumenta a variância genética total, embora diminua a diversidade alélica dentro de indivíduos. Entretanto, subdividindo a população em várias subpopulações por meio de amostragem aleatória, a variância genética dentro de subpopulações equivará a $(1-F)$ vezes a variância genética original. E esta redução na variação genética é o que ocorre em populações com tamanho efetivo baixo (que são amostras de uma população base). Por sua vez, a variação genética entre subpopulações aumenta, podendo conduzir a uma diferenciação entre elas.

1.10.8. Expressões do N_e úteis ao melhoramento e conservação genética de espécies perenes

(a) Amostragem em populações naturais dióicas

Na amostragem em populações naturais, o número de machos (N_m) amostrados é desconhecido, mas pode-se supor $N_m = N_f k_f$, em que N_f é o número de matrizes amostradas e k_f é o número médio de indivíduos amostrados (plantados) por matriz. Neste caso, o número médio de descendentes por macho (k_m) é igual a 1 e, tomando-se o mesmo número de descendentes por matriz $\sigma_{k_f}^2 = \sigma_{k_m}^2 = 0$, tem-se:

$$N_e = \frac{4 N_m N_f}{N_m + N_f} = \frac{4 N_f k_f N_f}{N_f k_f + N_f} = \frac{4 N_f k_f}{k_f + 1} \approx 4 N_f \text{ (Resende et al., 1995a).}$$

Por esta expressão verifica-se que a instalação de bancos de germoplasma ex situ plantando 10, 20 e 30 indivíduos por família conduz a tamanhos efetivos da ordem de 90%, 95% e 96%, respectivamente, do N_e que seria obtido plantando infinitos indivíduos por família. Assim, o benefício de utilizar mais que 20 indivíduos por família é pequeno.

(b) Amostragem em populações naturais monóicas alógamas

$$N_e = \frac{4 N_f k_f}{k_f + 3} \text{ (Vencovsky, 1986)}$$

(c) Seleção em populações experimentais monóicas alógamas com variados números de indivíduos selecionados por família de meios irmãos

$$N_e = \frac{4 N_f k_f}{k_f + 3 + (\sigma_{k_f}^2 / k_f)} \text{ (Resende & Bertolucci, 1995) , em que } N_f, k_f \text{ e } \sigma_{k_f}^2 \text{ referem-se ao número de famílias selecionadas, número médio de indivíduos selecionados por família e variância do número de indivíduos selecionados por família.}$$

Casos particulares desta expressão são:

(c.1.) Seleção de um mesmo número de indivíduos por família (ou seleção entre famílias de tamanho constante ou seleção entre e dentro de progênies)

$$N_e = \frac{4 N_f k_f}{k_f + 3}$$

Esta expressão é também dada alternativamente por $N_e = \frac{N_f k_f}{(1 - r) + k_f r} = 1/(2F)$, em que $F = (1/(2N))[(1 - r) + k_f r]$, sendo $N = N_f k_f$ e $r = 0,25$ para meios irmãos.

(c.2.) Seleção dentro de famílias, mantendo apenas um indivíduo por família

$$N_e = N_f$$

(c.3.) Tamanho efetivo de uma família

$$N_e = \frac{4 k_f}{k_f + 3}$$

Este N_e coincide com o N_s (*status number*) de uma família de meios-irmãos. O conceito de N_s foi apresentado por Lindgren et al. (1996) e refere-se à metade do inverso da coancestria (θ) média em um grupo (de tamanho k_f) de indivíduos, incluindo a coancestria de um indivíduo com ele mesmo. Para família de meios-irmãos

$$N_s = \frac{1}{2\theta} = \frac{k_f}{0,25 k_f + 0,75} = \frac{4 k_f}{k_f + 3}, \text{ em que } \theta = (1/8)+(3/8)/k_f.$$

(d) Número efetivo de famílias selecionadas

$N_{ef} = (\sum k_{fi})^2 / \sum k_{fi}^2$, onde k_{fi} é o número de indivíduos selecionados na família i (Robertson, 1961).

(e) Diversidade genética

$$D = \frac{N_{ef}}{N_{fo}}, \quad (0 < D \leq 1), \text{ onde } N_{fo} \text{ é o número original de famílias.}$$

(f) Seleção em populações experimentais monóicas com sistema reprodutivo misto e variados números de indivíduos selecionados por família de polinização aberta

$$N_e = \frac{2(2-S) N_f k_f}{(1+S)^2 \frac{\sigma_{k_f}^2}{k_f} + (1+S)^2 k_f + (3-2S-S^2)}, \text{ em que S refere-se à taxa de autofecundação.}$$

(g) Seleção em populações experimentais autógamas com variados números de indivíduos selecionados por linhagem

$$N_e = \frac{2N_f k_f}{\frac{\sigma_{k_f}^2}{k_f} + 4k_f} = \frac{0,5N_f k_f}{\frac{\sigma_{k_f}^2}{k_f} + k_f}$$

Uma derivação completa das expressões aqui apresentadas é mostrada no capítulo 18, tópico 18.7, baseando-se no conceito do tamanho efetivo de variância.

1.11. Heterose

1.11.1. Conceitos

Ao contrário da endogamia, que se refere a cruzamentos entre indivíduos aparentados, na exogamia os cruzamentos ocorrem entre indivíduos (ou grupos

de indivíduos) que possuem coeficiente de parentesco menor que o médio da população. A exogamia ou cruzamento entre entidades geneticamente divergentes promove o aparecimento do fenômeno denominado heterose, o qual, em princípio, está associado com a heterozigose promovida pela exogamia.

A heterose ou vigor híbrido é geralmente definida como a superioridade da média dos filhos em relação à média dos pais. Neste sentido, só pode ser explicada geneticamente, se for considerada alguma dominância associada ao caráter em questão. A heterobeltiose, por sua vez, é um caso particular de heterose em que a média dos filhos supera a média do genitor superior.

Para determinados caracteres, os genótipos heterozigóticos dos filhos podem apresentar média inferior à média dos pais. Alguns autores referem-se a este fenômeno como “heterose negativa” e este só pode ser explicado por dominância negativa, ou seja, dominância no sentido de diminuir a expressão do caráter.

Em termos biométricos, Falconer (1989) relata que para um loco com dois alelos a heterose é dada por:

$$\begin{aligned} He &= M_{F1} - M_p \\ &= (p - p')^2 d, \text{ em que:} \end{aligned}$$

M_{F1} : média da 1ª geração filial descendente do cruzamento entre duas populações.

M_p : média das populações genitoras.

p e p' : freqüências dos alelos favoráveis nas duas populações.

d : valor genotípico do heterozigoto (associado ao efeito de dominância).

A partir desta expressão conclui-se:

- (i) a magnitude da heterose depende do grau de diversidade ($p - p'$) alélica entre as populações e do nível de dominância.
- (ii) a heterose só ocorre se houver dominância e diversidade alélica, simultaneamente.
- (iii) a heterose será máxima quando um alelo é fixado em uma população e outro alelo, na outra população.

Estas conclusões permitem inferir que incrementar, sob seleção recíproca, a endogamia em populações genitoras divergentes e, posteriormente, cruzá-las é um procedimento que permite maximizar a heterose em um caráter com dominância.

Ao nível dos vários (n) locos de um caráter quantitativo, a heterose pode ser expressa por:

$$H_e = \sum_{i=1}^n (p_i - p'_{i })^2 d_i$$

A partir desta expressão, conclui-se que:

- (i) Se alguns locos apresentam dominância positiva e outros dominância negativa, os efeitos podem ser cancelados e não haver manifestação de heterose. Assim, a heterose em um caráter quantitativo depende de dominância direcional e a ausência de heterose não é suficiente para inferir sobre ausência de dominância.
- (ii) A heterose é a soma dos efeitos heteróticos dos locos em heterozigose na geração filial.

Com base nestas conclusões, poder-se-ia inferir que a heterose seria tanto maior quanto maior fosse a diversidade alélica entre as populações. Entretanto, isto nem sempre é verificado na prática devido à ocorrência de interações epistáticas e combinações gênicas coadaptadas que são quebradas quando populações extremamente divergentes são cruzadas (Falconer, 1989).

A heterose é calculada por $H_e = (MC - MP)/MP$, em que MC é a média da progénie obtida por cruzamento e MP é a média das populações genitoras.

1.11.2. Tipos de heterose

Resultados de heterose em várias espécies têm evidenciado que o seu nível é inversamente proporcional à herdabilidade (no sentido restrito) do caráter. Assim, maior heterose tem sido verificada para características de produção e outras relacionadas com a reprodução, as quais apresentam herdabilidades baixas.

Em plantas, podem ser definidos três tipos de heterose (Lamkey & Edwards, 1999):

- **Baseline heterosis:** heterose básica verificada no âmbito intrapopulacional por meio da comparação entre a média do híbrido de duas linhagens extraídas da mesma população com a média destas linhagens. Neste caso, o vigor perdido durante a endogamia é restaurado no F_1 e, pode-se dizer que esta heterose é o efeito oposto ao da depressão endogâmica.
- **Heterose panmítica:** também denominada heterose funcional, verifica-se no cruzamento entre duas populações comparando-se a média da população híbrida com a média das duas populações genitoras per se. Esta é a heterose relevante para espécies perenes e é definida no nível interpopulacional e não no nível intrapopulacional, como o é, a depressão endogâmica.

Em espécies anuais, em que são utilizados híbridos de linhagens obtidas de populações divergentes, a heterose em relação à média dessas linhagens equivale à heterose panmítica mais a heterose básica. Neste caso, a heterose é função da divergência genética, dominância e depressão endogâmica. A heterose panmítica é função apenas da divergência genética e da dominância.

Capítulo 2

**Progresso Genético
e Métodos de Seleção**

2.1. Conceitos

O termo seleção é definido como a reprodução diferencial dos diferentes genótipos na natureza ou sob intervenção do homem. A seleção natural opera independentemente das ações humanas e se realiza por meio de diferenças em fertilidade e sobrevivência (viabilidade) das progêniens dos diferentes indivíduos, tendo como consequência a evolução orgânica. A seleção artificial é praticada pelo homem e é baseada em critérios definidos pelo próprio melhorista, sendo assim, fator primordial para o melhoramento genético.

Essencialmente, a seleção atua promovendo a alteração das freqüências alélicas nos locos que controlam o caráter sob seleção, conduzindo a alteração na média genotípica da população, na direção desejada. Em princípio, na prática da seleção, o melhorista atua em duas etapas básicas: (*i*) a predição do valor genético dos indivíduos; (*ii*) a decisão sobre a melhor forma de utilização dos indivíduos com os maiores valores genéticos preditos.

2.2. Valores Genéticos, Fenotípicos e Seleção

Considerando um caráter (tal qual a produção de frutos em cafeeiros da espécie robusta) que pode ser avaliado em vários estágios ou idades dos indivíduos, tem-se o seguinte modelo (ignorando a epistasia) para as observações fenotípicas em um dado estágio (Turner & Young, 1969):

$$y = \mu + a + d + e_p + e_t \text{, em que:}$$

y : valor fenotípico temporário.

μ : média genotípica ou média geral.

a : efeito aditivo.

d : efeito de dominância.

e_p : efeito de ambiente permanente.

e_t : efeito de ambiente temporário.

$\mu + a$: valor genético aditivo (é um parâmetro similar à capacidade geral de combinação, equivalendo a essa capacidade multiplicada por um regressor, que é função da herdabilidade, e somada à média geral).

$\mu + a + d = \mu + g$: valor genotípico.

$\mu + a + d + e_p$: valor fenotípico permanente.

Este modelo é adequado a um caráter avaliado em um só local (ambiente). Neste caso, o efeito da interação genótipo x ambiente encontra-se confundido ou implícito em g . Para um caráter avaliado em um só estágio, os componentes ambientais permanentes e temporários também ficam confundidos em $e = e_p + e_t$.

Ao praticar a seleção, o melhorista de plantas perenes pode ter como alvo dois tipos de população melhorada de referência: (*i*) uma formada por descen-

dentes dos indivíduos selecionados e (ii) outra formada pelos próprios indivíduos selecionados, propagados clonalmente. O primeiro tipo referenciado em (i) corresponde à população da geração seguinte à seleção e o segundo tipo (ii) corresponde à população na mesma geração e clonada.

Assim, é relevante a predição dos valores genéticos aditivos e genotípicos, para atender as populações-alvo relatadas em (i) e (ii), respectivamente. Isto porque somente os efeitos aditivos dos alelos são transmitidos à descendência via reprodução sexuada - população de referência (i) - e porque na propagação clonal dos indivíduos selecionados - população de referência (ii) -, capitalizam-se os efeitos aditivos e de dominância.

Os valores genéticos (VG) e valores genotípicos (VGp) são variáveis aleatórias desconhecidas e devem ser preditas a partir do conhecimento de y . De maneira genérica, o princípio da regressão linear deve ser empregado para previsão de VG e VGp, ou seja, a técnica de regressão permite inferir em que proporção y explica ou prediz VG e VGp. Assim:

$$VG = \mu + \beta_{ay} (y - \mu)$$

$VGp = \mu + \beta_{gy} (y - \mu)$, em que β_{ay} é o coeficiente de regressão do valor genético aditivo sobre o valor fenotípico y e β_{gy} é o coeficiente de regressão do valor genotípico sobre o valor fenotípico.

Pelo princípio da regressão são os seguintes os estimadores:

$$\hat{\beta}_{a,y} = \frac{COV[(\mu + a), y]}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_y^2} = h_a^2; \hat{\beta}_{g,y} = \frac{COV[(\mu + g), y]}{\sigma_y^2} = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_y^2} = h_g^2, \quad \text{em}$$

que:

h_a^2, h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

$\sigma_y^2 = Var(\mu + a + d + ep + et) = \sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2 + \sigma_{et}^2 = \sigma_a^2 + \sigma_d^2 + \sigma_e^2$, considerando covariância zero entre os efeitos em y (a covariância entre efeitos genéticos e ambientais pode ser assumida como nula desde que haja casualização na distribuição dos genótipos nos ambientes, ou seja, desde que não seja dado tratamento preferencial aos melhores genótipos).

$\sigma_a^2, \sigma_d^2, \sigma_{ep}^2$ e σ_{et}^2 : variâncias genética aditiva, de dominância, ambiental permanente e ambiental temporária, respectivamente.

Os estimadores de β apresentados são adequados à seleção individual e serão utilizados para conceituação e apresentação do tema progresso genético e seus componentes.

2.2.1. Inferências pontuais e intervalares sobre valores genéticos

Os valores genéticos são variáveis aleatórias não observáveis que devem ser preditas pelo melhorista a partir dos valores fenotípicos observáveis. Em termos da inferência estatística paramétrica, é conveniente ao melhorista trabalhar tanto com a predição pontual quanto com a predição intervalar. A predição pontual fornece os valores genéticos preditos, ao passo que a predição intervalar propicia os intervalos de confiança dos valores genéticos e dos ganhos genéticos. A inferência intervalar é mais adequada ao melhorista, pois permite maior segurança na recomendação de uso dos indivíduos. Os intervalos de confiança (funções de uma quantidade pivotal) permitem que as decisões quanto à seleção sejam baseadas, preferencialmente, nos limites inferiores dos intervalos de confiança dos valores genéticos preditos dos indivíduos. Mais detalhes sobre a predição intervalar são apresentados no tópico 2.3.4.

2.3. Equação Geral do Progresso Genético

Uma vez que os genitores (diploídes) passam para sua progênie uma amostra correspondendo à metade de seus alelos, o valor genético esperado de uma progênie equivale à média dos valores genéticos preditos de seus genitores masculino e feminino. Por este mesmo princípio, a média da população melhorada (geração descendente) equivalerá à média dos valores genéticos adicionais dos indivíduos selecionados, supondo-se a seleção de um mesmo número de machos e de fêmeas e contribuição eqüitativa de gametas para a geração descendente.

O ganho genético pode ser definido como a superioridade (diferença) da população melhorada em relação à população não melhorada, ou seja, $\mu_1 - \mu_0$. A média genotípica da população melhorada (μ_1) corresponde à média dos valores genéticos dos indivíduos selecionados ao passo que a média da população não melhorada (μ_0) corresponde à média dos valores genéticos de todos os indivíduos da população.

Assim, os ganhos genéticos (G_s) podem ser estimados fazendo-se:

$$\begin{aligned}\hat{G}_s &= \mu_1 - \mu_0 \\ &= \sum_{i=1}^N \hat{V}G_i / N - \mu_0 \\ &= \mu_0 + \sum_{i=1}^N [\beta(y_i - \mu_0)] / N - \mu_0 \\ &= \beta(\sum_{i=1}^N y_i / N - N\mu_0 / N) \\ &= \beta(\bar{y}_s - \mu_0)\end{aligned}$$

em que N é o número de indivíduos selecionados e \bar{y}_s é a média fenotípica dos indivíduos selecionados.

O termo $\bar{y}_s - \mu_0$ pode ser definido como o diferencial de seleção (ds) e β refere-se ao coeficiente de regressão dos valores genéticos aditivos ou valores genotípicos sobre o valor fenotípico. Assim, se o objetivo do melhorista consiste na seleção por valores genéticos aditivos, β refere-se à herdabilidade no sentido restrito. Se o objetivo é a seleção por valores genotípicos (propagação clonal dos indivíduos selecionados), β diz respeito à herdabilidade no sentido amplo.

Levando em conta a seleção massal (individual) de indivíduos visando ao melhoramento da descendência, $\hat{G}_s = d_s h_a^2$ é a expressão padrão para a estimativa de ganhos genéticos (Lush, 1937).

Considerando a distribuição normal [$y \sim N(\mu, \sigma^2)$] dos dados observados, uma outra expressão para o ganho genético pode ser derivada. Uma variável aleatória com média μ e variância σ^2 pode ser transformada em uma outra variável padronizada denominada z, a qual possui média zero e variância 1, por meio da subtração de cada realização da variável aleatória em relação a sua média e posterior divisão por seu desvio padrão. Assim, para a variável y tem-se:

$$z_i = (y_i - \mu_0) / \sigma_y$$

Ao selecionar os melhores indivíduos de uma população padronizada, seleciona-se uma fração p de indivíduos, os quais são superiores a determinado indivíduo com valor z'. Neste caso, z' é o ponto de truncamento. O valor médio da proporção p selecionada (\bar{z}_s) é obtido pela função distribuição acumulada da distribuição normal padronizada, por:

$$\bar{z}_s = \frac{1}{p} \cdot \frac{1}{\sqrt{2\pi}} \int_{z'}^{\infty} z e^{-\frac{z^2}{2}} dz = \frac{1}{p} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z'^2}{2}} = \frac{1}{p} f(z') = \frac{z}{p}$$

Assim, $\bar{z}_s = \frac{z}{p}$, em que z é a ordenada da curva normal padronizada no ponto de truncamento z'.

Com os dados padronizados, o diferencial de seleção é dado por $d_{sp} = \bar{z}_s - \bar{z} = \bar{z}_s - 0 = \frac{z}{p} = k$, onde k é o diferencial de seleção padronizado.

Assim, com os dados previamente padronizados, o ganho genético na escala padronizada (em unidades de desvio padrão), pode ser calculado por: $\hat{G}_{sp} = k \beta$.

Os valores calculados pela curva normal padronizada podem ser transformados para a escala original, fato que permite calcular o ganho na escala dos dados originais, porém utilizando-se o valor k . São as seguintes as igualdades:

$\bar{z}_s = \frac{\bar{y}_s - \mu_0}{\sigma_y} = \frac{d_s}{\sigma_y} = \frac{z}{p} = k$, em que \bar{y}_s é a média fenotípica dos indivíduos selecionados.

A Fig. 2.1, referente à curva normal padrão, ilustra a exposição realizada.

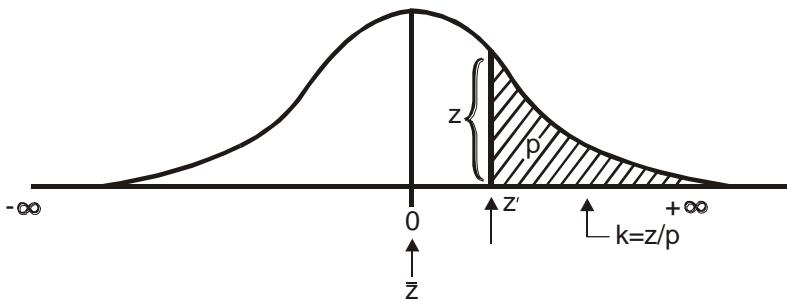


Fig. 2.1. Curva normal padrão, ilustrando a proporção (p) selecionada (área hachurada) e o diferencial de seleção padronizado (k), associados ao ponto de truncamento z' .

Assim, $d_s = \bar{y}_s - \mu_0 = k \sigma_y$, e, portanto, $\hat{G}_s = k \beta \sigma_y$, fornece o ganho na escala original.

Os valores de k podem ser obtidos diretamente por intermédio de z , o qual pode ser conseguido pela integração da função distribuição acumulada da curva normal padrão no intervalo que se estende do ponto de truncamento a infinito. k pode ser obtido de maneira mais simples, utilizando-se tabelas que fornecem o ponto de truncamento, a partir do conhecimento da proporção selecionada (área da curva normal padrão). Por exemplo, tendo-se $p = 0,10$, obtém-se $z' = 1,28$ e calcula-se $z = \frac{1}{\sqrt{2\pi}} e^{-\frac{z'^2}{2}} = 0,1758$ e $k = 0,1758/0,10 = 1,76$.

Os próprios valores de k são apresentados em função da proporção selecionada em várias obras (Falconer, 1989; Becker, 1984). Entretanto em situações de pequenas amostras, estas tabelas não são adequadas. Para amostras de tamanhos inferiores a 50, é recomendável a utilização da tabela apresentada por Becker (1984) ou o cômputo de um valor de k corrigido conforme Burrows (1972):

$$k_c = k - (1-p)/[2kp(N+1)], \text{ em que } k_c \text{ é o } k \text{ corrigido.}$$

Expressões equivalentes e alternativas a $\hat{G}_s = k \beta \sigma_y$ são:

$$\hat{G}_s = k r_{ya} \sigma_a \text{ para a seleção visando a propagação sexuada.}$$

$\hat{G}_s = k r_{yg} \sigma_g$ para a seleção visando a propagação assexuada.

Nas expressões apresentadas, os termos r_{ya} e r_{yg} , referem-se à acurácia ou à precisão na predição dos valores genéticos e genotípicos, respectivamente, utilizando a observação fenotípica como critério de seleção. Em termos genéticos, as estimativas r_{ya} e r_{yg} podem ser denotadas r_{aa} e r_{gg} , indicando que os valores genéticos e genotípicos foram preditos utilizando um índice baseado em valores fenotípicos.

Esta última (em termos da acurácia) forma de expressar o ganho é vantajosa, pois permite a comparação de métodos de seleção por meio de um só parâmetro, a acurácia. Isto porque σ_a e σ_g são propriedades da população e, portanto, constantes através dos métodos de seleção. O coeficiente k , também, em algumas situações pode ser considerado constante através dos métodos de seleção. σ_y e β por suas vezes variam através dos métodos.

Entretanto, a rigor, os métodos devem ser comparados via $G_s = k r_{aa}$ ou $G_s = k r_{gg}$, pois mesmo quando se seleciona o mesmo número de indivíduos de um mesmo número total de indivíduos, os valores de k associados aos diferentes métodos podem ser diferentes, devido à correlação intraclasse entre valores genéticos preditos (Resende, 1994).

Dessa forma, para efeito de comparação de métodos de seleção, deve-se introduzir uma correção ao cálculo de k , levando-se em consideração a correlação intraclasse (t_l) entre os valores genéticos preditos pelos diferentes métodos de seleção.

Empregando-se a expressão de Burrows (1972), Hill (1976) demonstrou que o k corrigido (k_c) pode ser obtido por:

$k_c = k - (1-p)/\{2 k p [N(1-t_l) + f t_l + 1]\}$, em que N é o tamanho da população e f é o número de famílias em avaliação. Esta expressão de k é baseada no número efetivo de indivíduos selecionados.

Uma expressão mais realista para o ganho genético, considerando k_c e o tamanho efetivo associado ao número de indivíduos selecionados, foi apresentada por Resende (1994): $G_s = k_c r_{aa} \sigma_a [1-1/(2N_e)]$, considerando a depressão endogâmica na mesma taxa do coeficiente de endogamia.

A expressão geral $G_s = k r_{aa} \sigma_a$ refere-se ao progresso genético em um ciclo de seleção. Uma expressão mais importante do ponto de vista prático refere-se ao progresso genético por ano ou unidade de tempo, a qual é dada por $G_s = (k r_{aa} \sigma_a)/L$, em que L diz respeito ao intervalo de gerações.

Por esta expressão, verifica-se que o aumento da eficiência dos programas de melhoramento genético ou aumento do ganho genético pode ser conseguido por meio de:

- Aumento da intensidade de seleção ou seleção de um menor número de indivíduos.

- Utilização de métodos mais acurados de predição de valores genéticos e seleção.
- Uso de germoplasma com maior variabilidade genética.
- Minimização do intervalo entre gerações.

2.3.1. Diferencial de seleção e intensidade de seleção

O diferencial de seleção na unidade de medição da característica equivale a $d_s = \bar{y}_s - \mu$ e em unidades de desvio padrão equivale a $k = \frac{\bar{y}_s - \mu}{\sigma_y}$. Assim, o

d_s pode ser expresso, alternativamente, por $\bar{y}_s - \mu$ ou por $k \sigma_y$. Sendo $k = z/p$, está explícita a dependência entre diferencial de seleção e proporção selecionada ou intensidade de seleção, ou seja, o diferencial de seleção será tanto maior quanto menor for a proporção selecionada.

Em espécies dióicas, as intensidades de seleção em cada sexo podem ser diferentes. Assim, o valor de k na expressão do ganho é dado por $k = \frac{1}{2} (k_m + k_f)$, em que k_m e k_f são os diferenciais de seleção padronizados associados à seleção de machos e fêmeas, respectivamente. Considerando diferentes intensidades de seleção e diferentes acurárias ($r_{\hat{a}am}$ e $r_{\hat{a}af}$) na seleção de machos e fêmeas, a expressão do ganho genético por ciclo de seleção é dada por $G_s = [(k_f r_{\hat{a}af} \sigma_a) + (k_m r_{\hat{a}am} \sigma_a)]/2$.

O diferencial de seleção apresentado na equação geral do progresso genético refere-se ao diferencial de seleção esperado, levando em conta que todos os genitores contribuem igualmente para a geração seguinte. A defasagem entre os diferenciais de seleção esperado e efetivo pode ser atribuída aos efeitos da seleção natural que atua, neste caso, principalmente sobre os caracteres reprodutivos.

2.3.2. Desvio padrão genético aditivo

A magnitude da variação genética aditiva é de grande relevância para o melhorista, pois está associada à magnitude da herdabilidade e da variabilidade fenotípica (e, conseqüentemente, do diferencial de seleção), visto que $\sigma_a^2 = h^2 \sigma_y^2$, e o desvio padrão genético aditivo é diretamente proporcional ao progresso genético. A variação genética não está muito sob o controle do melhorista, a não ser pela possibilidade de escolha ou geração de populações mais variáveis. Em geral, alguns caracteres apresentam pequena variação genética (como os caracteres reprodutivos) e outros apresentam considerável variação genética (como os caracteres associados à qualidade dos produtos).

2.3.3. Intervalo entre gerações

O intervalo entre gerações (L) é definido como o tempo decorrido entre o plantio dos candidatos à seleção e o plantio dos descendentes dos indivíduos selecionados. Assim, L depende da idade de seleção, do período necessário, após a seleção, para a produção de sementes e do período requerido, após a produção de sementes, para a produção de mudas e plantio. Dessa forma, esforços para minimizar cada um desses componentes muito contribuirão para o aumento da taxa de melhoramento genético das populações.

2.3.4. Acurácia, confiabilidade e variância do erro de predição

De maneira geral, os valores genéticos preditos não são iguais aos valores genéticos verdadeiros dos indivíduos. A proximidade entre estes dois valores pode ser avaliada com base na estatística denominada acurácia (Van Vleck et al., 1987), a qual refere-se à correlação entre os valores genéticos preditos e os valores genéticos verdadeiros dos indivíduos e é simbolizada por r_{aa} . Quanto maior a acurácia na avaliação de um indivíduo, maior é a confiança na avaliação e no valor genético predito deste indivíduo.

A acurácia seletiva depende da herdabilidade e repetibilidade do caráter, da quantidade e qualidade das informações e procedimentos utilizados na previsão dos valores genéticos. Como é uma medida que está associada à precisão na seleção, a acurácia é o principal elemento do progresso genético, em que o melhorista pode alterar visando maximizar o ganho genético.

A acurácia é, também, o parâmetro mais importante na comparação de métodos de seleção, pois quanto maior o seu valor, melhor preditor do valor genético verdadeiro é o método de seleção. O quadrado (r_{aa}^2) da acurácia é um coeficiente de determinação e é denominado confiabilidade (Goddard, 1992a).

A distância entre os valores genéticos preditos e os verdadeiros é função da quantidade ($1 - r_{aa}$) e a estatística denominada variância do erro de predição (PEV) do valor genético equivale a $\sigma_a^2(1 - r_{aa}^2)$ (Van Vleck et al., 1987). Assim, a confiabilidade expressa a proporção da variação (σ_a^2) dos valores genéticos verdadeiros que é explicada pelo índice ou método de seleção. A fração complementar ($1 - r_{aa}^2$) da determinação completa mede a fração de σ_a^2 não explicada pelo método.

A variância do erro de predição dos valores genéticos pode ser utilizada para construção de intervalos de confiança dos valores genéticos preditos, pela expressão $VG \pm t[(1 - r_{aa}^2) \sigma_a^2]^{1/2}$, em que t refere-se ao valor da distribuição t de Student associado a determinado nível de confiança ($t = 1,96$, para 95% de confiança). Por exemplo, tendo-se um valor genético predito equivalente a 3,5 kg de frutos, predito com acurácia igual a 0,70 e associado a uma variância genética aditiva de 0,10 kg², o intervalo de confiança do valor genético predito equivale a $3,5 \pm 1,96 [(1 - 0,70^2) 0,10]^{1/2} = 3,06$ a 3,94, com 95% de confiança.

A acurácia é um parâmetro parcialmente sob controle do melhorista já que seu valor máximo é limitado pela herdabilidade da característica, conforme pode ser verificado pelo estimador generalizado da acurácia:

$$\hat{r}_{aa} = \left(\frac{\sum \beta_i a_{i\alpha} h_a^2 \sigma_y^2}{h_a^2 \sigma_y^2} \right)^{1/2} = \left(\frac{\beta_1 a_{1\alpha} h_a^2 \sigma_y^2 + \beta_2 a_{2\alpha} h_a^2 \sigma_y^2 + \dots + \beta_n a_{n\alpha} h_a^2 \sigma_y^2}{h_a^2 \sigma_y^2} \right)^{1/2}, \quad \text{o}$$

qual simplifica-se para:

$$\hat{r}_{aa} = (\sum \beta_i a_{i\alpha})^{1/2} = (\beta_1 a_{1\alpha} + \beta_2 a_{2\alpha} + \dots + \beta_n a_{n\alpha})^{1/2}, \text{ em que:}$$

β_i : coeficiente de ponderação do componente i do índice de seleção.

$a_{i\alpha}$: correlação genética aditiva entre o componente i do índice e o valor genético aditivo.

De maneira genérica, o estimador da acurácia equivale a $\hat{r}_{aa} = \hat{\sigma}_{\hat{a}} / \hat{\sigma}_a$, em que $\sigma_{\hat{a}}$ e σ_a referem-se aos desvios padrões do índice de seleção e genético aditivo, respectivamente.

A variância do erro de predição ou PEV é dada por

$PEV = Var(\hat{a} - a) = \sigma_{\hat{a}}^2 - 2\sigma_{\hat{a}}\sigma_a + \sigma_a^2 = \sigma_a^2 - \hat{r}_{aa}^2 \sigma_a^2$. Da expressão da acurácia, obtém-se $\sigma_{\hat{a}} = r_{aa}\sigma_a$ e, portanto, $Var(\hat{a} - a) = \sigma_a^2 - r_{aa}^2\sigma_a^2 = \sigma_a^2(1 - r_{aa}^2)$. Esta expressão assume que os componentes de variância são conhecidos ou estimados com alta precisão.

A Tabela 2.1 apresenta os valores máximos de acurácia a serem obtidos com a seleção de indivíduos pela seleção combinada (procedimento ótimo de seleção para o caso balanceado e delineamento experimental inteiramente ao acaso com uma planta por parcela, ou blocos casualizados com uma planta por parcela e ajuste prévio para os efeitos de blocos, conforme Resende, 1991), usando o valor do indivíduo e também a média de n meios-irmãos.

Tabela 2.1. Valores da acurácia máxima (com $n \rightarrow \infty$) possível com a seleção combinada, em função da herdabilidade do caráter.

h^2	r_{aa} máx.
0,05	0,53
0,10	0,55
0,15	0,58
0,20	0,61
0,25	0,63

Continua...

Tabela 2.1. Continuação.

h^2	r_{aa} máx.
0,30	0,66
0,35	0,68
0,40	0,71
0,45	0,73
0,50	0,76
0,60	0,80
0,70	0,85
0,80	0,90
0,90	0,95

Os sumários de avaliação genética apresentam os efeitos genéticos adicionais preditos (\hat{a}) e a acurácia na avaliação de cada indivíduo. É importante ressaltar que os \hat{a} 's dos diferentes indivíduos podem ser comparados diretamente entre si, independentemente da acurácia. As diferentes quantidades e qualidades das informações dos indivíduos já foram consideradas no cálculo de \hat{a} (e também da acurácia). A acurácia simplesmente informa sobre o risco no uso de cada indivíduo e não tem a propriedade de informar se o indivíduo é bom ou não. Esta informação pode ser obtida exclusivamente com base em \hat{a} . Um indivíduo com alto \hat{a} e acurácia mais baixa do que a dos demais foi mais penalizado na avaliação genética do que os demais, ou seja, a média de sua progénie ou família foi mais descontada (forçada ou regredida em direção à média geral) do que as médias das progénies de outros indivíduos. Assim, tal indivíduo, embora apresente um intervalo de confiança mais amplo de seu valor genético predito, com menor limite inferior do intervalo de confiança, apresenta também maior limite superior do intervalo de confiança de seu valor genético predito, e pode, com alguma probabilidade razoável, de fato, ser melhor que os demais.

Em conclusão, a acurácia pode ser utilizada como um indicativo da intensidade de utilização dos indivíduos, sendo que os indivíduos com altos valores genéticos preditos e com acurácia mais baixa, devem ser utilizados com ressalvas.

A Tabela 2.2. apresenta um exemplo de um sumário de avaliação genética de clones. Os resultados (parciais) são relativos a clones de *Eucalyptus urophylla* avaliados na Venezuela.

Tabela 2.2. Sumário de avaliação genotípica de alguns clones de *Eucalyptus urophylla* na Venezuela. Caráter circunferência na altura do peito, avaliado em centímetros.

Número do clone	Número de rame- tes	Média	Efeito genotí- pico	Valor genotí- pico	Herdabili- dade da média	Acurácia	PEV	Intervalo de con- fiança (95%) do valor genotípico
1	26	38,88	6,27	38,48	0,90	0,95	1,02	36,5 – 40,5
2	41	38,32	5,89	38,05	0,94	0,97	0,61	36,5 – 39,6
3	23	35,04	2,50	34,71	0,89	0,95	1,12	32,6 – 36,8
4	24	34,46	2,33.	34,54	0,90	0,95	1,02	32,6 – 36,5
5	26	34,04	1,90	34,11	0,90	0,95	1,02	32,1 – 36,1
6	33	33,97	1,89	34,10	0,92	0,96	0,81	32,3 – 35,9
7	22	33,27	0,96	33,17	0,89	0,94	1,12	31,1 – 35,3
8	30	32,57	0,42	32,63	0,91	0,96	0,92	30,8 – 34,5
9	27	32,07	-0,14	32,07	0,91	0,95	0,92	30,2 – 34,0
10	37	31,97	-0,19	32,02	0,93	0,97	0,72	30,4 – 33,7

Média geral = 32,21 cm; h_g^2 ao nível de indivíduo = 0,27; $\sigma_g^2 = 10,20 \text{ cm}^2$; PEV: variância do erro de predição dos valores genotípicos.

Verifica-se pela Tabela 2.2 que o número de ramegos sobreviventes por clone foi suficiente para propiciar altos valores de acurácia seletiva. Os intervalos de confiança para os valores genotípicos permitem concluir, com 95% de confiança, que os clones 1 e 2 apresentam produtividades estatisticamente iguais. Neste caso, embora o clone 1 tenha apresentado maior valor genotípico predito, a menor PEV e maior acurácia associadas à predição do valor genotípico do clone 2 conduziram a intervalos de confiança praticamente idênticos para os dois clones.

2.4. Métodos de Seleção: Coeficientes de Ponderação das Informações Fenotípicas e Acurárias

A acurácia depende fundamentalmente das fontes de informação utilizadas para a predição dos valores genéticos e seleção. Para um único caráter, as principais fontes de informação ou unidades de seleção estão associadas aos procedimentos de seleção: individual, de famílias, de irmãos, dentro de família, entre e dentro de famílias, combinada, pela progénie e pela metodologia de modelos mistos (modelo BLUP individual). Neste capítulo serão apresentados estimadores associados à avaliação de famílias em experimentos com uma planta por parcela.

2.4.1. Seleção individual

Na seleção baseada em uma só avaliação no indivíduo, a acurácia equivale à raiz quadrada da herdabilidade. A seleção baseada na média de várias (m)

medidas no indivíduo apresenta como fator de ponderação da média fenotípica, a quantidade $\beta = \frac{m h_a^2}{1 + (m-1) \rho}$ e como acurácia $r_{aa} = \left(\frac{m h_a^2}{1 + (m-1) \rho} \right)^{1/2}$, em que ρ é o coeficiente de repetibilidade individual.

2.4.2. Seleção de famílias

Neste tipo de seleção, famílias (de meios-irmãos ou irmãos germanos) inteiras são selecionadas com base no comportamento médio de seus n componentes, e os próprios indivíduos avaliados são selecionados (seleção entre famílias sem o uso de sementes remanescentes). Para este caso, o fator de ponderação (herdabilidade da média de famílias) do efeito de família equivale a

$$\beta = \frac{[1 + (n-1) \rho_a] h_a^2}{1 + (n-1) [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]}, \text{ em que } \rho_a \text{ é a correlação genética entre}$$

os indivíduos do tipo de família considerada ($\rho_a = \frac{1}{2}$ e $\rho_a = \frac{1}{4}$, para irmãos germanos e meios-irmãos, respectivamente) e ρ_d é a correlação genética de dominância ($\rho_d = \frac{1}{4}$ e $\rho_d = 0$ para irmãos germanos e meios-irmãos, respectivamente). A acurácia é dada por:

$$r_{aa} = \left[\frac{1 + (n-1) \rho_a}{n} \frac{[1 + (n-1) \rho_a] h_a^2}{1 + (n-1) [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]} \right]^{1/2}$$

2.4.3. Seleção de irmãos

Um grupo de indivíduos pertencentes a uma família são selecionados com base no comportamento médio de seus irmãos (seleção entre famílias com o uso de sementes remanescentes). Neste caso, os indivíduos selecionados não são avaliados e, portanto, não contribuem para o cálculo da média da família. O

fator de ponderação equivale a $\beta = \frac{n \rho_a h_a^2}{1 + (n-1) [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]}$ e a acurácia equivale a:

$$r_{aa} = \left[\frac{n \rho_a^2 h_a^2}{1 + (n-1) [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]} \right]^{1/2}$$

Considerando m medições por indivíduo:

$$\beta = \frac{mn\rho_a h_a^2}{1 + (m-1) \rho + (n-1) m [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]}.$$

2.4.4. Seleção dentro de famílias

Neste tipo de seleção, os indivíduos são selecionados com base no desvio de seu valor fenotípico em relação à média de sua família, ou seja, não são eliminadas famílias. O fator de ponderação equivale a $\beta = \frac{(1-\rho_a) h_a^2}{(1-t)}$ e a acurácia é dada por:

$r_{aa} = \left[\frac{n-1}{n} \frac{(1-\rho_a)^2}{(1-t)} h_a^2 \right]^{1/2}$, em que $t = \rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)$ refere-se à correlação fenotípica intraclassé dentro de progênies.

A seleção dentro de progênies pode ser restrita a mesmo número de indivíduos selecionados por progénie ou irrestrita, ou seja, podem ser selecionados diferentes números de indivíduos por progénie (seleção massal na mistura de progênies, porém mantendo-se pelo menos um indivíduo selecionado por progénie). A seleção irrestrita permite maior intensidade efetiva de seleção.

2.4.5. Seleção com base na progénie

Nesta modalidade de seleção, os progenitores femininos ou masculinos (em geral, femininos) são selecionados com base no comportamento de suas progênies de meios irmãos. O coeficiente de ponderação para o desvio da média da progénie (MP) em relação à média geral (μ) é $\beta = \frac{2 n h_a^2}{4 + (n-1) h_a^2} = \frac{2 n}{n + (4 - h_a^2)/h_a^2}$, em que $(4 - h_a^2)/h_a^2$ é uma constante para um determinado caráter em uma população. À medida que n aumenta, β tende a 2.

A acurácia seletiva equivale a

$$r_{aa} = \left[(1/2) \frac{2 n}{n + (4 - h_a^2)/h_a^2} \right]^{1/2} = \left[\frac{n}{n + (4 - h_a^2)/h_a^2} \right]^{1/2}.$$

2.4.6. Seleção com base nos valores genéticos preditos dos genitores

Os ponderadores dos valores genéticos preditos dos genitores são β_1 e β_2 .

$$\beta_1 = 1/2$$

$$\beta_2 = 1/2$$

$$r_{aa} = [(1/4) (r_{aa}^2 + r_{am}^2)]^{1/2}$$

em que r_{aa} e r_{am} são as acurárias associadas à predição dos valores genéticos dos genitores feminino e masculino, respectivamente.

2.4.7. Seleção entre e dentro de famílias

Este é um tipo de seleção em dois estágios, em que, inicialmente, selecionam-se as melhores famílias e, em seguida, os melhores indivíduos dessas melhores famílias.

Os coeficientes de ponderação e acurárias em cada estágio equivalem àqueles descritos nos itens 2.4.2. e 2.4.4. Neste caso, o ganho genético por ciclo é dado por:

$G_s = k_e r_{aae} \sigma_a + k_d r_{aad} \sigma_a = \sigma_a (k_e r_{aae} + k_d r_{aad})$, onde k_e e k_d são os diferenciais de seleção padronizados para a seleção entre e dentro de famílias, respectivamente, e r_{aae} e r_{aad} são as acurárias associadas à seleção entre e dentro de famílias, respectivamente.

2.4.8. Seleção combinada

Nesta modalidade de seleção, as informações do indivíduo e de sua família são combinadas pela atribuição de pesos ótimos às duas fontes de informação. Neste caso, a seleção ótima é baseada em

$$I_{ij} = \beta_1 (\bar{Y}_{i\cdot} - \bar{Y}_{..}) + \beta_2 (Y_{ij} - \bar{Y}_{i\cdot}), \text{ em que:}$$

$$\beta_1 = \frac{[1 + (n-1) \rho_a] h_a^2}{1 + (n-1) \rho_a h_a^2},$$

considerando progênie de meios-irmãos.

$$\beta_2 = \frac{(1 - \rho_a) h_a^2}{(1-t)} = \frac{(1 - \rho_a) h_a^2}{1 - \rho_a h_a^2}$$

Y_{ij} : valor individual

$\bar{Y}_{i\cdot}$: média da família

$\bar{Y}_{..}$: média geral.

A acurácia seletiva é dada por:

$$r_{aa} = \left[\frac{[1 + (n-1) \rho_a]^2 h_a^2}{n [1 + (n-1) \rho_a h_a^2]} + \frac{n-1}{n} \frac{(1-\rho_a)^2}{(1-\rho_a h_a^2)} h_a^2 \right]^{1/2}$$

A seleção combinada foi apresentada primeiro por Lush (1945; 1947) e pode ser aplicada ao melhoramento de plantas perenes quando são utilizadas parcelas de uma planta no delineamento inteiramente ao acaso ou, em blocos ao acaso com prévio ajuste dos valores individuais para os efeitos de blocos. O ajuste é dado por $Y_{ij} - \bar{Y}_{.j} + \bar{Y}_{..}$, em que $\bar{Y}_{.j}$ é a média bloco (Resende, 1991). Com várias plantas por parcela, o procedimento ótimo de seleção inclui mais que duas fontes de informação e é denominado índice multiefeitos (Resende & Higa, 1994a), o qual será detalhado a partir do Capítulo 5.

2.4.9. Seleção combinada com base em avaliações repetidas em cada indivíduo da família de meios-irmãos

$\beta_1 = \frac{[1 + (n-1) \rho_a]/n}{[1 + (m-1) \rho + (n-1) m \rho_a h_a^2]/nm} h_a^2$: ponderador do efeito da média de família.

$\beta_2 = \frac{(1-\rho_a)}{[1 + (m-1) \rho - m \rho_a h_a^2]/m} h_a^2$: ponderador do efeito da média do indivíduo dentro de família.

$$r_{aa} = \left[\frac{[[1 + (n-1) \rho_a]/n]^2}{[1 + (m-1) \rho + (n-1) m \rho_a h_a^2]/nm} h_a^2 + \frac{\left[(1-\rho_a) \frac{n-1}{n}\right]^2}{[1 + (m-1) \rho - m \rho_a h_a^2][(n-1)/n]/m} h_a^2 \right]^{1/2}.$$

2.4.10. Seleção pela metodologia de modelos mistos: BLUP sob modelo de plantas individuais

O procedimento BLUP sob modelo individual usa todas as informações disponíveis, dos indivíduos e de seus ancestrais e progêneres, gerando valores genéticos preditos, de todos os indivíduos envolvidos na análise. Tal procedimento é apresentado detalhadamente a partir do Capítulo 6.

2.4.11. Predição de valores genotípicos por seleção combinada

Para a seleção e clonagem de indivíduos avaliados em testes de progêneres, torna-se importante a predição dos valores genotípicos e não apenas dos valores genéticos aditivos. Neste caso, o índice para a predição dos valores genotípicos é dado por:

$$I_{ij} = \beta_1(\bar{Y}_{i..} - \bar{Y}_{..}) + \beta_2(Y_{ij} - \bar{Y}_{i..}), \text{ em que:}$$

$$\beta_1 = \frac{[1 + (n-1) \rho_a] h_a^2 + [1 + (n-1) \rho_d] (h_g^2 - h_a^2)}{1 + (n-1) [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]};$$

$$\beta_2 = \frac{(1-\rho_a) h_a^2 + (1-\rho_d) (h_g^2 - h_a^2)}{[1 - \rho_a h_a^2 - \rho_d (h_g^2 - h_a^2)]}.$$

Os demais termos estão definidos no item 2.4.8. Para o caso balanceado, este índice equivale ao BLUP genotípico.

Preditores de valores genotípicos a partir de testes de progêneres com indivíduos repetidos clonalmente foram apresentados por Resende & Araújo (1993).

2.5. Comparação de Métodos de Seleção

Usando todas as informações disponíveis, o procedimento BLUP propicia o maior ganho genético possível em um ciclo de seleção e é, na pior das hipóteses, igual aos outros métodos. Dentre os demais, as superioridades relativas variam com as diferentes situações. No presente tópico, serão comparados alguns métodos que podem ser aplicados na situação em que se dispõe de informações em termos de indivíduos em uma população com estrutura de famílias de meios-irmãos. Para tanto, serão comparados os métodos: seleção individual, seleção de famílias, seleção dentro de famílias, seleção entre e dentro de famílias e seleção combinada.

Considere-se o método de seleção generalizado dado por:

$$I_{ij} = \beta_1(\bar{Y}_{i..} - \bar{Y}_{..}) + \beta_2(Y_{ij} - \bar{Y}_{i..})$$

em que:

I_{ij} : índice de seleção.

Tem-se:

- (a) seleção individual: $\beta_1 = \beta_2 = 1$;
- (b) seleção de famílias: $\beta_1 = 1$; $\beta_2 = 0$;
- (c) seleção dentro de famílias: $\beta_1 = 0$; $\beta_2 = 1$;
- (d) seleção combinada: $\beta_1 = \frac{(1-t)[1 + (n-1) \rho_a]}{(1-\rho_a)[1 + (n-1) t]}$; $\beta_2 = 1$, em que t é a correlação fenotípica intraclass, $t = \rho_a h_a^2$.
- (e) seleção entre e dentro de progêneres: $\beta_1 = 1$ e $\beta_2 = 0$ no primeiro estágio e

$\beta_1 = 0; \beta_2 = 1$ no segundo estágio. Verifica-se que com:

$$t = \rho_a \Rightarrow (d) = (a);$$

$t = 0$ e n grande, (d) tende a (b).

Sendo, $t = \rho_a h_a^2$, t tende a ρ_a , quando h_a^2 tende a 1, ou seja, com o aumento da herdabilidade, a vantagem da seleção combinada sobre a seleção individual decresce paulatinamente. Por outro lado, t tende a zero quando h_a^2 tende a zero e, neste caso com n grande, reduz-se a vantagem da seleção combinada sobre a seleção entre famílias.

Considerando a mesma intensidade de seleção, exceção feita à seleção entre e dentro de progêneres, os demais métodos podem ser comparados para cada nível de herdabilidade, apenas em função do parâmetro acurácia. A Tabela 2.3 apresenta a eficiência de três métodos de seleção, sobre a seleção massal.

Tabela 2.3. Eficiência relativa da seleção entre progêneres (E), dentro de progêneres (D) e seleção combinada (C) sobre a seleção massal, para diferentes valores de herdabilidade (h_a^2) e para $n = 60$ indivíduos por família.

h_a^2	E	D	C	C com $n \rightarrow \infty$	C com $n = 10$
0,05	1,54	0,75	1,71	2,36	1,21
0,10	1,29	0,75	1,50	1,75	1,18
0,20	1,02	0,76	1,28	1,36	1,12
0,30	0,87	0,77	1,17	1,20	1,09
0,40	0,77	0,78	1,12	1,11	1,06
0,50	0,70	0,80	1,06	1,07	1,04
0,60	0,65	0,81	1,03	1,04	1,02
0,70	0,60	0,82	1,02	1,02	1,01
1,00	0,51	0,86	1,00	1,00	1,00

Verifica-se que, com 60 indivíduos por família, a seleção entre famílias só é superior à seleção individual, quando a h_a^2 for menor que 0,20. Com o aumento de n , a seleção combinada torna-se ainda mais vantajosa sobre a seleção individual, principalmente para valores de h_a^2 menores que 0,50. Nas condições estudadas, a seleção dentro de progêneres é sempre inferior aos demais métodos de seleção.

Visto que a seleção combinada é a mais eficiente, resta compará-la com a seleção entre e dentro de progêneres. As expressões do progresso genético para as duas modalidades de seleção são:

$$G_s = k_i \sigma_a \left[\frac{[1 + (n-1) \rho_a]^2 h_a^2}{n[1 + (n-1) \rho_a h_a^2]} + \frac{n-1}{n} \frac{(1-\rho_a)^2}{(1-t)} h_a^2 \right]^{1/2} \text{ para a seleção combinada;}$$

$$G_s = \sigma_a \left[k_e \left[\frac{[1 + (n-1) \rho_a]^2 h_a^2}{n[1 + (n-1) \rho_a h_a^2]} \right]^{1/2} + k_d \left[\frac{n-1}{n} \frac{(1-\rho_a)^2}{(1-t)} h_a^2 \right]^{1/2} \right] \text{ para a seleção entre e dentro de progênies, em que:}$$

k_i , k_e e k_d : referem-se aos diferenciais de seleção padronizados para a seleção combinada, entre progênies e dentro de progênies, respectivamente.

Verifica-se que as duas expressões apresentam similaridades, diferindo, sobretudo, em termos dos diferenciais de seleção padronizados. A Tabela 2.4 apresenta os valores dos componentes da acurácia pelos dois métodos de seleção.

Tabela 2.4. Valores associados aos componentes da acurácia pelos métodos da seleção entre e dentro de famílias e seleção combinada, para diferentes valores de herdabilidade e de indivíduos (n) por família.

$h_a^2 \setminus n$	Acurácia para seleção entre progênies (E)		Acurácia para seleção dentro de progênies (D)		Acurácia para seleção combinada (C)		(E+D)-C		C-E	
	10	60	10	60	10	60	10	60	10 ou 60	
0,05	0,22	0,34	0,16	0,17	0,27	0,38	0,11	0,13	0,04	
0,10	0,29	0,41	0,23	0,24	0,37	0,47	0,15	0,17	0,05	
0,20	0,38	0,46	0,33	0,34	0,50	0,57	0,21	0,23	0,11	
0,30	0,43	0,48	0,41	0,42	0,59	0,64	0,25	0,26	0,16	
0,40	0,47	0,49	0,47	0,50	0,67	0,70	0,28	0,29	0,21	
0,50	0,50	0,50	0,54	0,56	0,74	0,75	0,30	0,31	0,25	

Os valores de acurácia para a seleção dentro de progênies e a diferença ($E + D - C$) entre o somatório dos valores de acurácia para seleção entre e dentro, e da acurácia para a seleção combinada, praticamente não variam com a variação do número de plantas por progênie (Tabela 2.4).

Dessa forma, a comparação entre a seleção entre e dentro de progênies e combinada pode ser realizada comparando-se as expressões da Tabela 2.5.

Tabela 2.5. Ganhos genéticos em unidades de desvio padrão genético aditivo para a seleção entre e dentro de progênies e seleção combinada.

h^2_a	Alternativa 1		Alternativa 2	
	Entre e dentro de progênies	Combinada	Entre e dentro de progênies	Combinada
0,05	$k_e r_{\text{aae}} + k_d r_{\text{aad}}$	$k_i r_{\text{aae}} + k_i r_{\text{aad}} - 0,13 k_i$	$k_e r_{\text{aae}} + k_d 0,17$	$k_i r_{\text{aae}} + k_i 0,04$
0,10	$k_e r_{\text{aae}} + k_d r_{\text{aad}}$	$k_i r_{\text{aae}} + k_i r_{\text{aad}} - 0,17 k_i$	$k_e r_{\text{aae}} + k_d 0,24$	$k_i r_{\text{aae}} + k_i 0,05$
0,20	$k_e r_{\text{aae}} + k_d r_{\text{aad}}$	$k_i r_{\text{aae}} + k_i r_{\text{aad}} - 0,23 k_i$	$k_e r_{\text{aae}} + k_d 0,34$	$k_i r_{\text{aae}} + k_i 0,11$
0,30	$k_e r_{\text{aae}} + k_d r_{\text{aad}}$	$k_i r_{\text{aae}} + k_i r_{\text{aad}} - 0,26 k_i$	$k_e r_{\text{aae}} + k_d 0,42$	$k_i r_{\text{aae}} + k_i 0,16$
0,40	$k_e r_{\text{aae}} + k_d r_{\text{aad}}$	$k_i r_{\text{aae}} + k_i r_{\text{aad}} - 0,29 k_i$	$k_e r_{\text{aae}} + k_d 0,50$	$k_i r_{\text{aae}} + k_i 0,21$
0,50	$k_e r_{\text{aae}} + k_d r_{\text{aad}}$	$k_i r_{\text{aae}} + k_i r_{\text{aad}} - 0,31 k_i$	$k_e r_{\text{aae}} + k_d 0,56$	$k_i r_{\text{aae}} + k_i 0,25$

$r_{\text{aae}}, r_{\text{aad}}$ – acurácia pela seleção entre progênies e dentro de progênies, respectivamente.

Com base na Alternativa 2 (Tabela 2.5), verifica-se que a vantagem da seleção combinada sobre aquela entre e dentro de progênies advém, principalmente, de uma maior intensidade de seleção, visto que k_i é sempre muito superior a k_e e k_d . Na Tabela 2.6, a seguir, faz-se uma comparação entre os dois métodos, considerando uma população composta por 100 famílias com 60 indivíduos cada uma.

Tabela 2.6. Comparação (em termos de ganhos genéticos em unidades de desvio padrão genético aditivo) entre a seleção entre e dentro de famílias (ED) e a seleção combinada (C) considerando a seleção de 10 indivíduos (1 indivíduo em 10 famílias na seleção entre e dentro de famílias) e 60 indivíduos (3 indivíduos em 20 famílias na seleção entre e dentro de famílias) em uma população composta por 100 famílias com 60 indivíduos cada uma.

h^2_a	Seleção de 10 indivíduos					Seleção de 60 indivíduos				
	ED	$k_i r_{\text{aae}} + k_i r_{\text{aad}}$	C	Dif.	E=C/ED	ED	$k_i r_{\text{aae}} + k_i r_{\text{aad}}$	C	Dif.	E=C/ED
0,05	1,00	1,61	1,20	0,41	1,20	0,83	1,36	1,02	0,34	1,23
0,10	1,30	2,05	1,52	0,53	1,17	1,07	1,74	1,29	0,45	1,21
0,20	1,63	2,52	1,80	0,72	1,10	1,34	2,14	1,53	0,61	1,14
0,30	1,86	2,84	2,02	0,82	1,09	1,54	2,40	1,71	0,69	1,11
0,40	2,07	3,12	2,21	0,91	1,07	1,72	2,64	1,87	0,77	1,09
0,50	2,23	3,34	2,36	0,98	1,06	1,85	2,83	2,00	0,83	1,08

E: Eficiência da seleção combinada sobre a seleção entre e dentro de progênies;

Dif = Diferença: $k_i r_{\text{aae}} + k_i r_{\text{aad}} - C$.

As superioridades da seleção combinada são maiores para valores baixos de herdabilidade e para intensidades de seleção mais baixas (que pode significar também menor tamanho da população experimental e menor número de famílias avaliadas).

A Fig. 2.2 ilustra as principais características da seleção combinada em comparação com a seleção entre e dentro de progênies, dentro de progênies e massal.

Ordem dos melhores indivíduos dentro de progênies

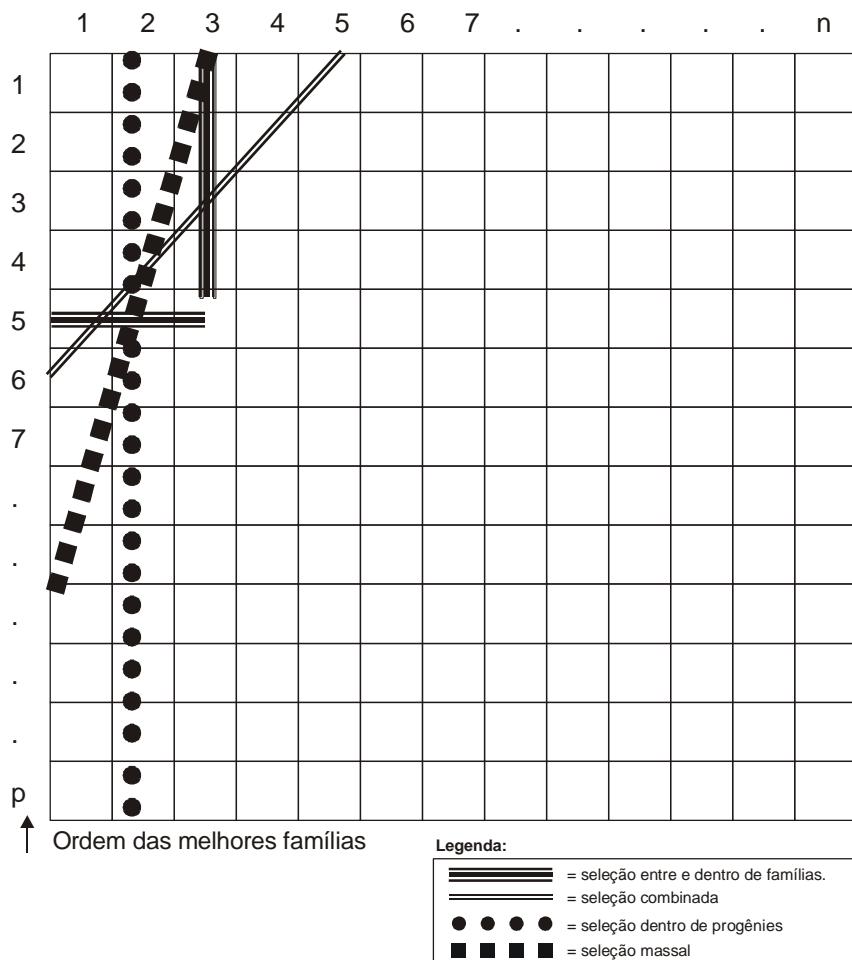


Fig. 2.2. Comparação de várias modalidades de seleção, em que os indivíduos selecionados pelos diferentes métodos situam-se nas áreas à esquerda das respectivas faixas.

Com base na Fig. 2.2, verifica-se que:

- A seleção entre e dentro de famílias recombinha os melhores indivíduos das melhores famílias ao passo que a seleção combinada recombinha também indivíduos excelentes de famílias médias e indivíduos médios de famílias excepcionais, eliminando indivíduos médios de famílias médias.
- A seleção combinada propicia maior intensidade de seleção, pois toma a área do triângulo na área total do retângulo, ao passo que a seleção entre e dentro de progêneres seleciona linhas no total de linhas e, posteriormente, colunas no total de colunas.
- A seleção combinada tende a propiciar menor tamanho efetivo populacional (N_e), pois existe uma variância do número (n) de indivíduos selecionados por família, ao passo que na seleção entre e dentro $\sigma_n^2 = 0$. Isto pode ser afirmado com base na expressão $N_e = 4N / (\sigma_n^2 + 2)$, segundo Falconer (1989), em que N refere-se ao número total de indivíduos selecionados.
- A seleção massal é intermediária entre a seleção dentro de famílias e seleção combinada, em termos de tamanho efetivo e ganho genético, propiciando N_e maior que a seleção combinada e menor que a seleção dentro de famílias, e ganho maior que a seleção dentro de progêneres e menor que a combinada.
- A seleção dentro de famílias propicia o maior N_e dentre todos os métodos comparados, pois preserva o número inicial de famílias e $\sigma_n^2 = 0$.

2.6. Correlação Intraclasse entre Valores Genéticos Preditos

É comum, após a predição dos valores genéticos dos indivíduos, verificar que existe uma certa correlação entre os valores genéticos preditos associados a indivíduos de uma mesma família, ou parentados em algum grau. Esta correlação intraclasse ocorre devido à covariância genética entre os indivíduos parentados e, em alguns casos, por causa da covariância ambiental entre indivíduos desenvolvidos em um mesmo microambiente. Esta correlação é importante do ponto de vista prático porque reduz o diferencial de seleção padronizado e o tamanho efetivo populacional e, consequentemente, afeta o progresso genético no curto e longo prazos.

Estes coeficientes de correlação intraclasse variam com os métodos de seleção e os métodos que enfatizam a informação de família tendem a produzir valores genéticos preditos mais correlacionados e menores intensidades de seleção realizadas do que os métodos que atribuem pequeno peso à informação de família.

Os estimadores para os coeficientes de correlação intraclasse entre valores genéticos preditos são dados por:

$$\hat{t}_i = \frac{\text{COV} (I_{ij}, I_{ij'})}{\sigma_{I_{ij}} \sigma_{I_{ij'}}}, \text{ em que:}$$

$COV (I_{ij}, I_{ij'})$: covariância entre os índices ou valores genéticos preditos dos indivíduos j e j' pertencentes à família i.

$\sigma_{I_{ij}}^2 = \sigma_{I_{ij'}}^2$: variância do índice de seleção ou dos valores genéticos preditos.

Os estimadores t_i podem ser derivados empregando-se propriedades de esperança matemática associados a variáveis aleatórias. Assim, tem-se, por exemplo, para a seleção massal:

$$\begin{aligned} COV (I_{ij}, I_{ij'}) &= COV [h_a^2(Y_{ij} - \bar{Y} \cdot \cdot), h_a^2(Y_{ij'} - \bar{Y} \cdot \cdot)] \\ &= h_a^4 COV [(Y_{ij} - \bar{Y} \cdot \cdot), (Y_{ij'} - \bar{Y} \cdot \cdot)] \\ &= h_a^4 (1/4) \sigma_a^2 \end{aligned}$$

$$\begin{aligned} \sigma_{I_{ij}}^2 &= Var [h_a^2(Y_{ij} - \bar{Y} \cdot \cdot)] \\ &= h_a^4 Var [(Y_{ij} - \bar{Y} \cdot \cdot)] \\ &= h_a^4 \sigma_y^2 \end{aligned}$$

$$\hat{t}_I = \frac{COV (I_{ij}, I_{ij'})}{\sigma_{I_{ij}}^2} = \frac{(1/4) \sigma_a^2}{\sigma_y^2} = \rho_a h_a^2$$

Para a seleção combinada, considerando um grande número de indivíduos por família de meios-irmãos, $t_I = \frac{1-0,25 h_a^2}{1+2 h_a^2}$ (Robertson, 1961). Com base

nesta expressão, os valores de t_I são: 0,90; 0,81; 0,68; 0,58 e 0,50 para herdabilidades de 0,05; 0,10; 0,20; 0,30 e 0,40, respectivamente. Estes valores revelam que a correlação é problemática, sobretudo, para valores baixos de h_a^2 . Estimadores de t_I para outros métodos de seleção, como o índice multifeitos foram apresentados por Resende (1994).

2.7. Tendência Genética e Herdabilidade Realizada

A tendência genética refere-se a alterações na composição genética das populações, em termos de freqüências alélicas e genotípicas, como resultado da atuação de fatores evolutivos tais como a seleção, a deriva genética, a migração e a mutação. Caracteriza-se, então, como uma mudança no estado de equilíbrio das populações.

Em programas de melhoramento genético, a tendência genética é resultante principalmente da seleção. Assim, neste caso, equivale ao ganho genético realizado, o qual pode ser medido, comparando-se, em condições ambientais idênticas, populações de diferentes gerações. Em termos genéticos, a tendência genética (TG) é dada por $TG = \mu_{G1} - \mu_{G0}$, onde μ_{G1} e μ_{G0} referem-se às médias da geração atual e original, respectivamente. Na prática, a tendência genética pode ser computada via metodologia de modelos mistos, em que a média dos

valores genéticos preditos dos indivíduos pertencentes a determinada geração fornece uma estimativa da média da geração.

De posse da estimativa do ganho genético realizado (TG) e conhecendo-se o diferencial de seleção (ds) praticado, pode ser estimada a herdabilidade realizada dada por $h_a^2 = TG/ds$.

2.8. Efeitos da Seleção na Variância, nos Parâmetros Genéticos e no Tamanho Efetivo

A seleção altera os componentes de médias e variâncias dos valores genéticos e fenotípicos por meio de: mudanças nas freqüências alélicas, desequilíbrio em fase gamética, endogamia e deriva genética. Em termos estatísticos, produz mudanças nos primeiros (médias) e segundos (variâncias) momentos das variáveis aleatórias de uma população base, causadas em função de um processo de amostragem não aleatória dessa população.

Sob um modelo poligênico puramente aditivo, mesmo para populações infinitas, a seleção direcional modifica a variância genética pela indução de desequilíbrio de ligação (Lush, 1945; Bulmer, 1971). Para populações de tamanho finito, a seleção afeta a estrutura de família, aumentando a perda de variação genética via endogamia (Lush, 1946; Robertson, 1961).

A seleção reduz a variação genética de duas maneiras (Bulmer, 1971; 1980). A primeira é pela criação do desequilíbrio em fase gamética. Este fator é o mais imediato e, inicialmente, o de maior efeito. Neste caso, a seleção de genitores reduz a variância entre eles e, consequentemente, reduz a covariância entre os indivíduos da progénie, de forma que a herdabilidade calculada tende a ser inferior à verdadeira.

A segunda maneira pela qual a variância é reduzida é resultante das próprias alterações nas freqüências alélicas nos locos que afetam o caráter, em função da seleção. Entretanto, as freqüências alélicas mudam muito devagar quando o número de locos é grande e, consequentemente, os efeitos de cada loco na expressão do caráter são muito pequenos. Assim, na prática, os efeitos decorrentes de alterações nas freqüências alélicas podem ser negligenciados, exceto para a seleção no longo prazo.

Considerando apenas o efeito da seleção de genitores, as variâncias fenotípicas (σ_y^2) e genética (σ_a^2) da população original são reduzidas por um fator k^* de forma que as novas variâncias (σ_y^{2*}) e (σ_a^{2*}) são dadas por:

$$\sigma_a^{2*} = (1 - h_a^2 k^*) \sigma_a^2 ;$$

$$\sigma_y^{2*} = (1 - k^*) \sigma_y^2 .$$

O fator k^* é dependente da intensidade de seleção e equivale a $k^* = k(k - t)$, em que k é o diferencial de seleção padronizado (intensidade de seleção) e t é o desvio do ponto de truncamento em relação à média geral.

A fração $-h_a^2 k^* \sigma_a^2$ refere-se à covariância devido ao desequilíbrio em fase gamética e, sendo negativo, reduz a variação genética.

Na geração descendente dos genitores selecionados, somente a variância genética aditiva entre progênies de irmãos germanos é reduzida, equivalendo a $(1/2) (1 - h_a^2 k^*) \sigma_a^2$. A variação genética aditiva total (entre + dentro de progênies) equivale a $(1/2) (1 - h_a^2 k^*) \sigma_a^2 + (1/2) \sigma_a^2 = [1 - (1/2) h_a^2 k^*] \sigma_a^2$. A variância fenotípica por sua vez equivale a $[1 - (1/2) h_a^4 k^*] \sigma_y^2$, de forma que a herdabilidade equivalerá a $h_a^2 = \frac{1 - (1/2) h_a^2 k^*}{1 - (1/2) h_a^4 k^*} h_a^2$, portanto inferior à da geração genitora. Com baixas herdabilidades, entretanto, a redução do ganho genético na geração subsequente a da seleção é muito pequena. Com $h_a^2 = 0,10$, a redução no ganho genético varia de 4% a 3% para proporções de seleção variando de 5% a 50%.

Com novas seleções, novos desequilíbrios serão gerados e a variação genética aditiva após g gerações de seleção será dada por $\sigma_{a(g+1)}^2 = (1/2) [1 - h_{a(g)}^2 k^*] \sigma_{a(g)}^2 + (1/2) \sigma_a^2$. Mas, a redução em σ_a^2 não ocorre indefinidamente, pois um equilíbrio é atingido quando o desequilíbrio gerado pela seleção é contrabalançado pela perda de desequilíbrio, causada por recombinação. Em ausência de ligação, três a quatro gerações de seleção conduzem ao equilíbrio (Falconer, 1989). Assim, não é conflitante o fato de a seleção não conduzir a redução de ganhos em algumas gerações após a seleção, em alguns experimentos de campo.

Em algumas situações de campo, a redução na variação fenotípica, em decorrência da seleção, também não ocorre, e isto pode ser explicado por dois fatores: efeito de escala ou correlação média-variância e aumento da variação ambiental contrabalançando a redução de σ_a^2 , já que homozigotos podem apresentar maior variação ambiental associada a eles (Falconer, 1989).

O efeito da seleção sobre o tamanho efetivo (N_e) foi estudado por Robertson (1961), que apresentou a expressão: $N/N_e = 1 + 2 k^2 h_a^2 [1 - h_a^2 k (k-t)]$. Assim, o N_e em relação à população de tamanho N, se reduz progressivamente com o aumento da herdabilidade e da intensidade de seleção.

2.9. Variância da Resposta à Seleção

Considerando a expressão da variância do erro de predição dos valores genéticos dada por $Var(\hat{a} - a) = PEV = (1 - r_{aa}^2) \sigma_a^2$ e a seleção de N indivíduos, com acurárias idênticas, e em igual proporção de sexo, a variância do ganho genético (valor genético médio dos N indivíduos selecionados) em uma geração é dada por (Resende et al., 1995b):

$V(G_s) = \frac{1}{N} [(1 - r_{aa}^2) \sigma_a^2]$, considerando os componentes de variância conhecidos ou estimados com precisão, tal como na filosofia do BLUP.

A variância (σ_R^2) da resposta (R) à seleção após várias gerações de melhoramento pode ser derivada a partir do modelo de Nicholas (1980), que considera R após g gerações de melhoramento:

$R = \bar{x}_g - \bar{x}_o$, em que \bar{x}_g é a média da população após g gerações de melhoramento e \bar{x}_o é a média da população original não melhorada.

$\sigma_R^2 = \sigma_o^2 + \frac{1}{M} \sigma_y^2$, em que σ_o^2 é a variância devida a deriva ou oscilação genética, M é o número total de indivíduos avaliados em cada geração e σ_y^2 é a variância fenotípica.

Conforme Hill (1977), $\sigma_o^2 = g \sigma_a^2 / N_e$, em que N_e é o tamanho efetivo populacional.

Ignorando-se o termo σ_y^2 / M (para M suficientemente grande), têm-se:

$$\sigma_R^2 = g \sigma_a^2 / N_e \text{ para } g \text{ gerações de seleção;}$$

$$\sigma_{R^*}^2 = \sigma_a^2 / N_e \text{ para uma geração de seleção.}$$

Verifica-se que a expressão de $\sigma_{R^*}^2$ é similar à de $V(G_s)$, sendo $\sigma_{R^*}^2$ função do número efetivo e $V(G_s)$ função do número físico de indivíduos selecionados. A expressão de $V(G_s)$ é corrigida (via $1 - r_{aa}^2$) em função da confiabilidade (r_{aa}^2) da seleção, o que não o é $\sigma_{R^*}^2$. Dessa forma, esta última é conservadora (superestimada) em relação à primeira e poderá ser usada com mais segurança.

Intervalos de confiança do ganho genético podem ser construídos por $G_s \pm t [V(G_s)]^{1/2}$ ou $G_s \pm t \sigma_{R^*}$, sendo que esta última expressão penaliza mais o limite inferior do intervalo de confiança.

A intensidade ótima de seleção na população de produção (visando a ganho genético no curto prazo) depende, sobretudo, da variância do ganho genético com seleção. Quanto menor o número de indivíduos selecionados, maior é a intensidade de seleção e, consequentemente, maior é também a variância do ganho genético, fato que é indesejável por significar menor probabilidade de realização do ganho genético predito.

A variância do ganho genético para os sistemas de propagação sexuada ou assexuada, respectivamente, é dada por:

$$V(G_s) = \frac{1}{N} (1 - r_{aa}^2) \sigma_a^2 \text{ ou}$$

$$V(G_s) = \frac{1}{N} (1 - r_{gg}^2) \sigma_g^2$$

Assim, esta variância é tanto menor quanto maiores forem a acurácia e o número de indivíduos selecionados.

No sistema de propagação clonal, geralmente os indivíduos são selecionados a partir de testes clonais, nos quais r_{gg} tende a 1 e, portanto, $V(G_S)$ tende a 0. Dessa forma, um número muito pequeno ($N \rightarrow 1$) de clones pode ser selecionado.

No sistema de propagação sexuada, a seleção de indivíduos está associada a uma acurácia de média magnitude (geralmente de 0,40 a 0,70) e, portanto, o N ideal deve ser determinado criteriosamente. Deve-se, para tanto, considerar a maximização do limite inferior (LI) do intervalo de confiança do ganho genético predito, dado por $LI = kr_{aa} - t \sqrt{(1/N)(1-r_{aa}^2)}$ (desconsiderando-se σ_a

que é constante), em que t é a estatística de Student, equivalendo a 1,96 para um intervalo construído com 95% de confiança. A referida expressão pode ser rescrita como:

$$LI = (z/p) r_{aa} - 1,95 \sqrt{(1/N)(1-r_{aa}^2)} = (z/N/nfb) r_{aa} - 1,95 \sqrt{(1/N)(1-r_{aa}^2)} = (znfb/N) r_{aa} - 1,95 \sqrt{(1/N)(1-r_{aa}^2)},$$

em que z é a ordenada da curva normal padrão no ponto de truncamento, p é a fração selecionada e n , f e b referem-se aos números de indivíduos por parcela, de famílias e de blocos, respectivamente. Com base nesta última expressão, e considerando $nfb = 10.000$ como ideal, pode ser estabelecida uma tabela (Tabela 2.7) que indica o N ótimo para cada grau de controle genético (herdabilidade no sentido restrito = h_a^2).

Tabela 2.7. Número de indivíduos (N) a serem selecionados para composição de um pomar de sementes clonal, a partir de experimentos com f famílias e nb indivíduos por família, em função da herdabilidade no sentido restrito (h_a^2) associada ao caráter sob seleção.

h_a^2	95% de $r_{aa\max}$	nb^*	F	N	Ganho genético em uni- dades de desvio padrão genético aditivo
0,10	0,5225	200	50	25	1,2907
0,15	0,5510	120	85	25	1,3865
0,20	0,5795	85	120	20	1,4798
0,25	0,5985	50	200	20	1,5461
0,30	0,6270	35	285	15	1,6551

* valor necessário para atingir 95% da acurácia máxima possível.

Verifica-se que o número ótimo de indivíduos a serem incluídos no pomar de sementes varia de 15 a 25 e, logicamente, quanto maior a herdabilidade maior será o ganho genético obtido (Tabela 2.7).

Em espécies dióicas, com proporção de sexo 1:1 e com machos e fêmeas avaliados em um mesmo experimento, devem ser utilizados no pomar $(1/2)N$ indivíduos masculinos e $(1/2)N$ indivíduos femininos. Por outro lado, se a proporção de sexo difere de 1:1, estes valores devem se recalculados. Os valores ótimos de N devem ser determinados pela combinação que maximiza $[(k_m r_{aa} + k_f r_{aa}) / 2] - t \sqrt{(1/N_e)(1-r_{aa}^2)}$ (Resende et al., 1997a).

Capítulo 3

**Seleção Indireta e Uso
de Caracteres Auxiliares**

3.1. Resposta Correlacionada

Em presença de correlação genética de alta magnitude, a alteração em um caráter, via seleção, promove alterações significativas em outros caracteres correlacionados. Estas alterações são denominadas respostas correlacionadas ou ganho genético indireto.

O ganho genético indireto em uma característica y , em função da seleção na característica x é dado por:

$$\begin{aligned} RC(y/x) &= \beta_{ayx} G_{sx} = \frac{\text{cov}_{a(x,y)}}{\sigma_{ax}^2} k h_x^2 \sigma_x \\ &= \frac{\text{cov}_{a(x,y)}}{\sigma_{ax} \sigma_{ay}} k \frac{\sigma_{ax}}{\sigma_x} \sigma_{ax} \frac{\sigma_{ay}}{\sigma_{ay}} \\ &= k \frac{\text{cov}_{a(x,y)}}{\sigma_{ax} \sigma_{ay}} \frac{\sigma_{ax}}{\sigma_x} \sigma_{ay} \\ &= k r_{a(x,y)} h_x \sigma_{ay} \frac{\sigma_y}{\sigma_y} \\ &= k r_{a(x,y)} h_x h_y \sigma_y \end{aligned}$$

em que:

β_{ayx} : regressão genética aditiva da característica y em relação à característica x .

G_{sx} : ganho genético direto no caráter x .

$r_{a(x,y)}$: correlação genética aditiva entre os caracteres x e y .

$r_{(x,y)}$: correlação fenotípica entre os caracteres x e y .

h_x^2 : herdabilidade do caráter x .

k : diferencial de seleção padronizado.

σ_{ax}^2 : variância genética aditiva do caráter x .

σ_x : desvio padrão fenotípico do caráter x .

A resposta correlacionada pode ser expressa também em função do diferencial de seleção correlacionado ou indireto ($dscy$) em y , via seleção em x , pela expressão:

$$RC(y/x) = dscy \cdot h_y^2$$

Outra forma de estimar a resposta correlacionada é através do produto do diferencial de seleção do caráter x (dsx) por uma herdabilidade indireta (h_c^2) ou correlacionada do caráter y , pela expressão:

$$RC(y/x) = dsx \cdot h_c^2 = k \sigma_x h_c^2$$

3.2. Co-herdabilidade e Coeficiente de Predição Genética

A herdabilidade indireta ou correlacionada mencionada no tópico anterior é denominada co-herdabilidade (Baker, 1985) e é dada por:

$$h_c^2 = \frac{\text{cov}_{a(x,y)}}{\sigma_x^2}$$

A resposta correlacionada pode ser predita, também, utilizando-se o diferencial de seleção truncado do próprio caráter em que se quer medi-la. Neste caso, o diferencial de seleção do caráter y (dsy) deve ser multiplicado pelo parâmetro denominado coeficiente de predição genética (Baradat, 1982; Gallais, 1989).

Tomando-se a expressão para a resposta correlacionada $RC(y/x) = k r_{a(x,y)} h_x h_y \sigma_y$ e sendo $dsy = k \sigma_y$, tem-se então que o coeficiente de predição genética equivale a $\psi_{x,y} = r_{a(x,y)} h_x h_y$ e a resposta correlacionada pode ser predita por $RC(y/x) = k \sigma_y \psi_{x,y} = dsy \psi_{x,y}$. Assim, $\psi_{x,y}$ equivale ao produto das acurárias (associadas à seleção massal) e da correlação genética entre os dois caracteres.

O parâmetro $\psi_{x,y}$ é mais adequado que a co-herdabilidade para avaliar a eficiência da seleção indireta em relação à seleção direta. Esta eficiência pode ser avaliada de modo exato por $\psi_{x,y} / h_y^2$, ou seja, basta comparar diretamente $\psi_{x,y}$ com h_y^2 . Se $\psi_{x,y} > h_y^2$, a seleção indireta será mais vantajosa que a seleção direta.

3.3. Seleção Indireta

Conforme visto em tópico anterior, a expressão mais simplificada para a resposta correlacionada equivale a $RC(y/x) = k h_x r_{a(x,y)} \sigma_{ay}$, ou seja, esta resposta depende da intensidade e acurácia da seleção em x e também da correlação genética entre os caracteres x e y. Verifica-se, também, que quando a variação genética (ou herdabilidade) em qualquer dos caracteres é muito pequena, a resposta correlacionada também o será.

A eficiência da seleção indireta pode ser expressa por:

$$\frac{RC_{y/x}}{G_{sy}} = \frac{k h_x r_{a(x,y)} \sigma_{ay}}{k h_y \sigma_{ay}} = \frac{h_x r_{a(x,y)}}{h_y} = \frac{\psi_{x,y}}{h_y^2}$$

Verifica-se por essa expressão que a seleção indireta só será vantajosa sobre a seleção direta quando $h_x r_{a(x,y)} > h_y$, ou seja, quando o produto da acurácia de x pela correlação genética entre x e y for maior que a acurácia de y. É importante notar que mesmo quando a correlação genética for igual a um, ainda é necessário que h_x^2 seja maior que h_y^2 , para que a seleção indireta seja mais

efetiva. Não sendo $r_{a(x,y)}$ igual a 1, esta deve ser de alta magnitude e $h_x^2 > h_y^2$ para que a seleção indireta possa ser mais efetiva.

É interessante notar que a acurácia da seleção indireta equivale a $r_{a(x,y)} h_x$ e que a eficiência da seleção indireta equivale também à razão entre as acuráncias indireta e direta. Ademais, a seleção via um caráter z será mais eficiente que a seleção indireta via um caráter x se $r_{a(y,z)} h_z > r_{a(y,x)} h_x$.

A predição indireta do valor genético de um indivíduo para um caráter y, pelo uso do valor genético predito (\hat{a}_x) para um caráter x é dada por

$$\hat{a}_y = (\text{cov}_{a(x,y)} / \sigma_{a_x}^2) \hat{a}_x = [r_{a(x,y)} (h_y / h_x) (\sigma_y / \sigma_x)] \hat{a}_x = \beta_{ayx} \hat{a}_x$$

3.4. Seleção Utilizando Caracteres Auxiliares ao Melhoramento

Outro sistema de seleção visando ao melhoramento de um único caráter é obtido por meio da utilização simultânea do caráter objetivo da seleção e de outros caracteres auxiliares, correlacionados àquele. Esse sistema é denominado seleção multivariada e baseia-se em um índice da seguinte forma:

$I = \sum_{i=1}^n b_i x_i$, em que x_i refere-se aos caracteres objetivo e auxiliares e b_i equivale ao coeficiente de ponderação associado ao caráter i.

Os pesos b_i são determinados por

$\tilde{b} = P^{-1} \tilde{g}$, em que:

\tilde{b} : vetor dos coeficientes de ponderação do índice.

P : matriz de covariância fenotípica envolvendo os caracteres incluídos no índice.

\tilde{g} : vetor de covariância genética entre os caracteres incluídos no índice e o objetivo da seleção.

Alternativamente \tilde{b} pode ser calculado por:

$\tilde{b} = P^{-1} G \tilde{w}$, em que:

G : matriz de covariância genética entre os caracteres incluídos no índice.

\tilde{w} : vetor contendo 1 para o caráter objetivo e 0 para os demais caracteres.

3.5. Herdabilidade Multivariada e Coeficiente de Predição Genética Multivariado

O parâmetro herdabilidade multivariada (h_m^2) é útil na comparação entre os métodos seleção direta, seleção indireta e seleção multivariada. A herdabilidade multivariada foi assim denominada por Gallais (1989) e refere-se à “herdabilidade” do índice. Provavelmente, o autor emitiu o conceito de herdabilidade

como equivalente a um coeficiente de determinação ou quadrado da acurácia seletiva (tal como ocorre no caso da seleção individual univariada) e não como regressão, pois no caso de índice de seleção, o coeficiente de herdabilidade como regressão equívale a 1. A h_m^2 é dada por:

$$h_m^2 = r_{la}^2 = \frac{(\text{cov}(I, a))^2}{\sigma_I^2 \sigma_a^2} = \frac{(b' g)^2}{(b' Pb) \sigma_a^2}, \text{ em que } g \text{ é o vetor coluna da matriz } G, \text{ correspondendo ao caráter objetivo da seleção e } r_{la} \text{ é a acurácia da seleção multivariada visando a ganho em um caráter } y.$$

No caso da seleção indireta multivariada, o conceito de h_m^2 pode também ser aplicado na obtenção da resposta correlacionada no caráter y pela seleção bivariada em x e y ($RC_{(y/x,y)}$), por meio da seguinte derivação:

$$RC(y/x, y) = k r_{la} \sigma_{ay}$$

$$= k r_{la} \sigma_{ay} \frac{\sigma_y}{\sigma_y}$$

$$= k \sigma_y r_{la} h_y$$

$$= k \sigma_y h_m h_y$$

Verifica-se, por esta expressão, que o coeficiente de predição genética multivariado equívale a $\psi_m = r_{la} h_y = h_m h_y$, ou seja, ao produto das acurárias das seleções multivariada ($r_{la} = h_m$) e direta (h_y). Observa-se, também, que o método multivariado é melhor se $r_{la} h_y > h_y^2$, ou seja $h_m > h_y$, que implica também $h_m^2 > h_y^2$, uma vez que a eficiência da seleção multivariada é dada por $Ef(y/x, y) = \psi_m / h_y^2 = h_m h_y / h_y^2 = h_m / h_y$. Em termos genéricos, $r_{la} = \sigma_I / \sigma_a$ e o coeficiente de predição genética multivariado é dado por $\psi_m = \sigma_I / \sigma_a h_y = \sigma_I / \sigma_y$.

Para a seleção multivariada, o ganho em precisão quando se usam caracteres auxiliares pode ser expresso em termos do número de repetições (b) a mais na avaliação dos genótipos, que seria necessário para se atingir a mesma precisão. Para se verificar isto basta igualar a herdabilidade multivariada à herdabilidade ao nível de médias de genótipos (Gallais, 1989):

$$h_m^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2 / b} \Rightarrow b = \frac{h_m^2}{1 - h_m^2} \frac{\sigma_e^2}{\sigma_g^2} = \frac{h_m^2}{h^2} \frac{1 - h^2}{1 - h_m^2}, \text{ em que } \sigma_g^2 \text{ pode ser equivalente à variação genotípica (seleção clonal) ou } (1/4) \sigma_a^2 \text{ (seleção entre famílias).}$$

Como exemplo, se $h_m^2 = 0,18$ e $h^2 = 0,10$, $b = 1,98 \approx 2$ e o ganho em precisão quando se usa a seleção multivariada é equivalente a 1 repetição.

Entretanto, parece mais adequado expressar b em função do ψ_m já que este multiplica diretamente o diferencial de seleção do caráter objetivo (y).

$$\text{Assim, } b = \frac{\psi_m}{h^2} \cdot \frac{1-h^2}{1-\psi_m}.$$

Especificamente no caso bivariado, empregando-se os caracteres objetivo da seleção (y) e um caráter auxiliar (x), tem-se que:

$$h_m^2 = r_{la}^2 = h_y^2 \left[1 + \frac{((h_x r_{a(x,y)} / h_y) - r_{(x,y)})^2}{1 - r_{(x,y)}^2} \right].$$

Assim, neste caso $h_m^2 > h_y^2$, ou seja, a herdabilidade bivariada é sempre superior à univariada (só será igual se $r_{a(x,y)} h_x / h_y$ for igual a $r_{(x,y)}$). A fração $[(h_x r_{a(x,y)} / h_y) - r_{(x,y)}]^2 / (1 - r_{(x,y)}^2)$ h_y^2 fornece o acréscimo na herdabilidade devido à utilização do caráter auxiliar. Esse acréscimo pode também ser obtido fazendo-se $h_m h_y - h_y^2$, ou seja, $\psi_m - h_y^2$.

Para este caso, os coeficientes b_1 e b_2 (que ponderam os valores fenotípicos dos caracteres), associados a y e x , respectivamente, são dados por

$$\begin{bmatrix} b_y \\ b_x \end{bmatrix} = \begin{bmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{yx} & \sigma_x^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{gy}^2 \\ \sigma_{gx}^2 \end{bmatrix} = \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix} \begin{bmatrix} \sigma_{gy}^2 \\ \sigma_{gx}^2 \end{bmatrix}, \text{ em que:}$$

$$C^{11} = \frac{\sigma_x^2}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2}; \quad C^{12} = \frac{-\sigma_{yx}}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2}; \quad C^{22} = \frac{\sigma_y^2}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2}.$$

Assim,

$$b_y = C^{11} \sigma_{gy}^2 + C^{12} \sigma_{gx}^2 = \frac{\sigma_x^2 \sigma_{gy}^2 - \sigma_{yx} \sigma_{gx}^2}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2}$$

$$b_x = C^{21} \sigma_{gy}^2 + C^{22} \sigma_{gx}^2 = \frac{\sigma_y^2 \sigma_{gy}^2 - \sigma_{yx} \sigma_{gx}^2}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2}$$

Fazendo-se $b_y = 1$, tem-se $b_x = \frac{\sigma_y^2 \sigma_{gx}^2 - \sigma_{yx} \sigma_{gy}^2}{\sigma_x^2 \sigma_y^2 - \sigma_{yx} \sigma_{gy}^2}$ (Searle, 1965), em que:

σ_y^2 e σ_x^2 : variâncias fenotípicas dos caracteres y e x , respectivamente.

σ_{gy}^2 e σ_{gx}^2 : covariâncias genética e fenotípica, respectivamente, entre os caracteres y e x .

σ_{gy}^2 : variância genética do caráter y .

Sendo $b_y = 1$, b_x pode também ser expresso em termos das correlações genéticas (r_a) e fenotípica (r) e dos coeficientes de herdabilidade dos caracteres 1 (h_y^2) e 2 (h_x^2):

$$b_x = \frac{\sigma_y}{\sigma_x} \frac{r_a h_x - rh_y}{h_y - rr_a h_x}$$

Quanto aos valores assumidos por b_x , pode-se dizer:

- Se $r_a h_x/h_y$ for igual a r , b_x será igual a zero e, portanto, o uso do caráter auxiliar não trará benefício algum. Entretanto, se as estimativas dos componentes de variância não forem fidedignas, o uso deste caráter auxiliar poderá ser pior do que o emprego apenas do caráter objetivo.
- se b_x assumir um valor negativo, o uso do caráter auxiliar será vantajoso, mas no sentido de indicar uma influência do ambiente sobre o caráter y , b_x negativo ocorre quando a correlação ambiental é muito maior que a correlação genética entre os dois caracteres.
- Com correlação genética igual a zero, $b_x = -\sigma_{yx}/\sigma_x^2$.

A eficiência ou ganho em acurácia com a seleção bivariada usando caracteres auxiliares é dada pela razão entre as acurárias bivariada e univariada, ou seja:

$$Ef = \frac{h_m}{h_y} = \left[1 + \frac{((h_x r_{axy}/h_y) - r_{xy})^2}{1 - r_{xy}^2} \right]^{1/2}$$

A eficiência da seleção usando caracteres auxiliares sobre a seleção indireta é dada por:

$$Ef = \frac{h_y}{h_x r_{axy}} \left[1 + \frac{(h_x r_{axy}/h_y - r_{xy})^2}{1 - r_{xy}^2} \right]^{1/2}$$

3.6. Índice de Seleção Biológico e Herdabilidade Generalizada

Uma extensão da seleção baseada em caracteres auxiliares (quando o caráter objetivo recebe peso 1 e os caracteres auxiliares recebem peso 0), refere-se à situação em que os valores genéticos de todos os caracteres são preditos simultaneamente usando caracteres auxiliares.

Neste caso, os coeficientes de ponderação são função apenas dos parâmetros genéticos, dispensando-se as informações econômicas. Assim, os índices gerados são de natureza puramente biológica e existirá um índice para cada caráter considerado como objetivo da seleção.

O termo “herdabilidade generalizada” foi introduzido por Gallais (1973) no contexto da seleção multivariada. Desejando-se predizer os valores genéticos (g) de vários caracteres simultaneamente por meio de seus valores fenotípicos (w), são os seguintes os estimadores:

$$\hat{g} = \beta' w \\ = GP^{-1} w$$

Assim, a herdabilidade multivariada equivale à função $\beta' = GP^{-1}$.

No caso de apenas dois caracteres, têm-se:

$$\begin{aligned}\beta' &= \begin{bmatrix} \beta_{yy} & \beta_{yx} \\ \beta_{xy} & \beta_{xx} \end{bmatrix} = GP^{-1} \\ P^{-1} &= \begin{bmatrix} \sigma_x^2 & -\sigma_{yx} \\ -\sigma_{yx} & \sigma_y^2 \end{bmatrix} / [\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2] \\ \beta_{yy} &= \frac{\sigma_{g_y}^2 \sigma_x^2 - \sigma_{yx} \sigma_{g_{yx}}}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2} = \frac{h_y^2 - r r_a h_y h_x}{1 - r^2} \\ \beta_{yx} &= \frac{\sigma_{g_{yx}} \sigma_y^2 - \sigma_{yx} \sigma_{g_y}^2}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2} = \frac{(\sigma_{g_y} / \sigma_x)(r_a h_x - r h_y)}{1 - r^2} \\ \beta_{xy} &= \frac{\sigma_{g_y g_x} \sigma_x^2 - \sigma_{yx} \sigma_{g_x}^2}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2} = \frac{\sigma_{g_x} / \sigma_y (r_a h_y - r h_x)}{1 - r^2} \\ \beta_{xx} &= \frac{\sigma_{g_x}^2 \sigma_y^2 - \sigma_{g_{yx}} \sigma_{yx}^2}{\sigma_y^2 \sigma_x^2 - (\sigma_{yx})^2} = \frac{h_x^2 - r r_a h_y h_x}{1 - r^2}, \text{ em que:}\end{aligned}$$

$\sigma_{g_{yx}}$ e σ_{yx} : covariâncias genética e fenotípica entre os caracteres y e x, respectivamente.

$\sigma_{g_y}^2$ e σ_y^2 : variâncias genética e fenotípica para o caráter y, respectivamente.

h_y^2 : herdabilidade do caráter y.

Assim:

$$g = \begin{bmatrix} \beta_{yy} & \beta_{yx} \\ \beta_{xy} & \beta_{xx} \end{bmatrix} \begin{bmatrix} w_y \\ w_x \end{bmatrix}$$

Com correlações genotípicas (r_a) e fenotípicas (r) nulas ou negligenciadas:

$\beta_{yy} = h_y^2$; $\beta_{yx} = \beta_{xy} = 0$; $\beta_{xx} = h_x^2$, conforme tratado no caso univariado.

O conceito de herdabilidade generalizada pode ser aplicado também na seleção pelo método dos níveis independentes de eliminação (Capítulo 4), quando a seleção não é baseada em um índice, mas o ganho total em um caráter é função dos ganhos em cada um dos estágios: ganho direto em um dos estágios (quando a seleção baseia-se no próprio caráter) e ganho indireto no outro estágio (quando a seleção para o outro caráter é efetuada). Turner & Young (1969) apresenta estimadores adequados a esta situação. Cotterill &

James (1981) consideraram, também, a questão da seleção pelo método dos níveis independentes em dois estágios.

3.7. Seleção em Múltiplos Estágios

A seleção em múltiplos estágios envolve, em geral, a eliminação de indivíduos inferiores em uma idade precoce e, em seguida, a seleção (entre os não eliminados previamente) dos indivíduos superiores na idade adulta.

Neste caso, o ganho genético total (G_{ST}) é dado por $G_{ST} = G_{S(A/J)} + G_{S(A/A)}$, em que:

$G_{s(A/J)} = k_J \sigma_{a(J,A)} / \sigma_{FJ}$: ganho na idade adulta com seleção no 1º estágio (juvenil).

$G_{s(A/A)} = K_A^* \sigma_{aA}^{2*} / \sigma_{FA}^*$: ganho na idade adulta com seleção no 2º estágio (adulto).

k_J : diferencial de seleção padronizado, na idade juvenil.

k_A^* : diferencial de seleção padronizado, na idade adulta.

σ_{aA}^{2*} : variância genética aditiva na idade adulta, ajustada para o efeito da seleção realizada no 1º estágio.

σ_{FA}^* : desvio padrão fenotípico na idade adulta, ajustado para o efeito da seleção realizada no 1º estágio.

$\sigma_{a(J,A)}$: covariância genética aditiva entre os estágios juvenil e adulto.

σ_{FJ} : desvio padrão fenotípico na idade juvenil.

Considerando as variáveis z_i , z_j e z_k e suas covariâncias σ_{ij} , σ_{ik} e σ_{jk} , os componentes de variância são ajustados para os efeitos da seleção na variável z_i , por meio da expressão genérica: $\sigma_{jk}^* = \sigma_{jk} - \sigma_{ij} \sigma_{ik} \sigma_{ii}^{-1} k_j (k_j - t_j)$, expressão esta construída a partir de Cochran (1951) e Cunningham (1975). A variável k_J ($k_J - t_J$) é puramente dependente da intensidade de seleção em z_i e foi tabulada por Finney (1956). A partir desta expressão genérica tem-se, por exemplo, a variância genética aditiva ajustada para uma seleção prévia realizada na mesma população e considerando o caráter como o mesmo ($r_{aJA} = 1$) de uma idade para outra:

$\sigma_{aA}^{2*} = \sigma_{aA}^2 - [(\sigma_{aA}^2)^2 k_J (k_J - t_J)] / \sigma_{FJ}^2 = \sigma_{aA}^2 [1 - h^2 k_J (k_J - t_J)]$, em que:

t_J : ponto de truncamento da curva normal padronizada, associado à seleção prévia.

h^2 : herdabilidade associada à seleção.

σ_{aA}^2 : variância aditiva na idade adulta, antes da seleção.

A fórmula do ganho total pode ser expressa alternativamente por:

$$\begin{aligned} G_{ST} &= k_J h_J h_A r_{aJA} \sigma_{FA} + k_A^* h_A^{2*} \sigma_{FA}^* \\ &= k_J h_J h_A r_{aJA} \sigma_{FA} + k_A^* \frac{1 - r_{aJA}^2 h_J^2 k^*}{(1 - r_{FA}^2 k^*)^{1/2}} h_A^2 \sigma_{FA} \end{aligned}$$

Essa igualdade advém dos seguintes resultados:

$$h_A^{2*} = h_A^2 \left(\frac{1 - r_{a_{JA}}^2 h_J^2 k^*}{1 - r_{FA}^2 k^*} \right); \quad \sigma_{FA}^{2*} = (1 - r_{FA}^2 k^*) \sigma_{FA}^2; \quad k^* = k_J (k_J - t_J); \quad \sigma_{aA}^{2*} = (1 - r_{a_{JA}}^2 h_J^2 k^*) \sigma_{aA}^2,$$

em que:

σ_{FA}^2 : variância fenotípica na idade adulta, não ajustada para o efeito da seleção realizada no 1º estágio.

h_A^2 : herdabilidade na idade adulta, não ajustada para o efeito da seleção.

h_A^{2*} : herdabilidade na idade adulta, ajustada para o efeito da seleção.

h_J^2 : herdabilidade na idade juvenil.

$r_{a_{JA}}$: correlação genética aditiva entre os estágios juvenil e adulto.

r_{FA} : correlação fenotípica entre os estágios juvenil e adulto.

Considere os seguintes parâmetros genéticos:

$h_J^2 = 0,25$: herdabilidade na idade juvenil.

$h_A^2 = 0,20$: herdabilidade na idade adulta, não ajustada para o efeito da seleção.

$r_{a_{JA}} = 0,80$: correlação genética aditiva entre os estágios juvenil e adulto.

$r_{FA} = 0,65$: correlação fenotípica entre os estágios juvenil e adulto.

$\sigma_{FA}^2 = 2,5$: variância fenotípica na idade adulta, não ajustada para o efeito da seleção realizada no 1º estágio.

Considerando uma intensidade de seleção de 10% no primeiro estágio e também no segundo estágio, tem-se $k_J = k_A^* = 1,755$, $t_J = 1,282$ e $k^* = k_J (k_J - t_J) = 1,755(1,755 - 1,282) = 0,8301$.

Assim, o ganho genético total com a seleção nos dois estágios equivale a:

$$G_{ST} = k_J h_J h_A r_{a_{JA}} \sigma_{FA} + k_A^* \frac{1 - r_{a_{JA}}^2 h_J^2 k^*}{(1 - r_{FA}^2 k^*)^{1/2}} h_A^2 \sigma_{FA}$$

$$G_{ST} = 1,755 (0,25)^{1/2} (0,20)^{1/2} 0,80 (2,5)^{1/2} + 1,755 \frac{1 - (0,80)^2 0,25 0,8301}{(1 - (0,65)^2 0,8301)^{1/2}} 0,20 (2,5)^{1/2} = 1,0937$$

Exercícios referentes aos procedimentos abordados neste Capítulo são apresentados ao final do capítulo 4.

Capítulo 5

**Avaliação Genética
pela Metodologia Tradicional**

Neste capítulo são considerados aspectos relativos à avaliação genética completa de indivíduos, a partir de um experimento balanceado no delineamento de blocos ao acaso, com várias plantas por parcela.

5.1. Teste de Progênie de Meios-irmãos

5.1.1. Estimativas de parâmetros genéticos e fenotípicos

Considere o experimento com $p=62$ progênies de meios-irmãos de acerola, avaliadas no delineamento de blocos ao acaso, com $b=3$ repetições e $n=4$ plantas por parcela, conduzido pela Embrapa Agroindustrial Tropical, em Fortaleza - CE. A análise de variância para os caracteres peso de frutos (PF) e diâmetro da copa (DC), revelou os seguintes resultados:

Fonte de variação	QM	$\frac{PF}{QM}$	$\frac{DC}{QM}$	E (QM)
Bloco	Q_1	385,006	0,41845	-
Progênie	Q_2	67,3101**	0,17912**	$\sigma_d^2 / n + \sigma_e^2 + b\sigma_p^2$
Resíduo	Q_3	37,3276	0,10106	$\sigma_d^2 / n + \sigma_e^2$
Dentro de parcela	Q_4	106,2997	0,35484	σ_d^2
Média do caráter		21,4815	3,2957	

Com base nos resultados da análise de variância, têm-se as seguintes estimativas, conforme Vencovsky & Barriga (1992):

Peso de Frutas	Diâmetro da copa
$\hat{\sigma}_p^2 = \frac{Q_2 - Q_3}{b} = 9,9942$	$\hat{\sigma}_p^2 = \frac{Q_2 - Q_3}{b} = 0,0260$
$\hat{\sigma}_e^2 = Q_3 - Q_4/n = 10,7527$	$\hat{\sigma}_e^2 = Q_3 - Q_4/n = 0,0123$
$\hat{\sigma}_d^2 = Q_4 = 106,2997$	$\hat{\sigma}_a^2 = 4\hat{\sigma}_p^2 = 0,1040$
$\hat{\sigma}_a^2 = 4\hat{\sigma}_p^2 = 39,9768$	$\hat{\sigma}_d^2 = Q_4 = 0,3548$
$\hat{h}_a^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_d^2} = 0,3147$	$\hat{h}_a^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_d^2} = 0,2646$
$\hat{\sigma}_f^2 = \hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_d^2 = 127,0466$	$\hat{\sigma}_f^2 = \hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_d^2 = 0,3931$
$\hat{c}^2 = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_F^2} = 0,0846$	$\hat{c}^2 = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_F^2} = 0,0313$

em que:

$\hat{\sigma}_p^2$: estimativa da variância genética entre progênies;

$\hat{\sigma}_e^2$: estimativa da variância entre parcelas;

$\hat{\sigma}_d^2$: estimativa da variância dentro de parcelas;

$\hat{\sigma}_a^2$: estimativa da variância genética aditiva;

$\hat{\sigma}_f^2$: estimativa da variância fenotípica ao nível de indivíduos;

\hat{h}_a^2 : estimativa da herdabilidade individual no sentido restrito, no bloco;

\hat{c}^2 : estimativa da correlação devida ao ambiente comum da parcela.

A análise de covariância entre os caracteres PF e DC apresentou os seguintes resultados:

Fonte de covariância	PM	PM	E (PM)
Bloco	P ₁	3,2976	-
Progénie	P ₂	0,01334	$\text{cov}_d/n + \text{cov}_e + b \text{cov}_p$ $\text{cov}_d/n + \text{cov}_e$
Resíduo	P ₃	0,44663	
Dentro de Parcela	P ₄	1,5876	cov_d

Com base nos resultados da análise de covariância, são apresentadas as seguintes estimativas, conforme Vencovsky & Barriga (1992):

$\hat{COV}_p = \frac{P_2 - P_3}{b} = -0,1444$: estimativa da covariância genética entre os caracteres PF e DC, ao nível do efeito de progénie;

$\hat{COV}_e = P_3 - P_4/n = 0,0497$: estimativa da covariância ambiental entre os caracteres PF e DC, ao nível entre parcelas;

$\hat{COV}_d = P_4 = 1,5876$: estimativa da covariância fenotípica entre os caracteres PF e DC, dentro de parcelas.

$\hat{r}_a = \frac{\hat{COV}_p}{\hat{\sigma}_{p_{PF}} \hat{\sigma}_{p_{DC}}} = \frac{-0,1444}{[9,9942,0,02602]^{1/2}} = -0,2833$: estimativa da correlação genética aditiva entre os caracteres PF e DC.

$$\hat{r}_f = \frac{\hat{COV}_p + \hat{COV}_e + \hat{COV}_d}{\hat{\sigma}_{f_{PF}} \hat{\sigma}_{f_{DC}}} = \frac{1,4929}{[127,0466 \cdot 0,3931]^{1/2}} = 0,2112 : \text{estimativa da correlação fenotípica entre os caracteres PF e DC.}$$

Com base nessas estimativas, três procedimentos de seleção podem ser comparados, visando ao melhoramento do caráter peso de frutos: seleção direta em PF; seleção indireta em DC; seleção bivariada usando PF e DC (caráter x) como caráter auxiliar. As eficiências relativas podem ser computadas empregando-se os multiplicadores do diferencial de seleção padronizado direto do caráter PF. As expressões para o ganho em PF (caráter y) são:

$$\hat{G}_{s_{y/y}} = \hat{h}_{ay}^2 k \hat{\sigma}_{f_y} : \text{estimativa do ganho genético direto no caráter y;}$$

$$\hat{G}_{s_{y/x}} = \hat{\psi}_x k \hat{\sigma}_{f_y} : \text{estimativa do ganho genético indireto no caráter y via seleção em X;}$$

$$\hat{G}_{s_{y/(x/y)}} = \hat{\psi}_{x,y} k \hat{\sigma}_{f_y} = \text{estimativa do ganho genético em y por seleção bivariada em x e y.}$$

Têm-se as seguintes estimativas:

$$\hat{h}_{ay}^2 = 0,3147 : \text{estimativa da herdabilidade individual do caráter y (PF);}$$

$$\hat{h}_{ax}^2 = 0,2647 : \text{estimativa da herdabilidade individual do caráter x (DC);}$$

$$\hat{\psi}_x = \hat{h}_{ax} \hat{h}_{ay} = -0,0817 : \text{estimativa do coeficiente de predição genética univariado em x (DC);}$$

$$\hat{\psi}_{x,y} = \hat{h}_y^2 \left[1 + \frac{(\hat{h}_{ax} \hat{r}_a / \hat{h}_{ay}) - \hat{r}_f}{1 - \hat{r}_f^2} \right]^{1/2} = 0,3493 : \text{coeficiente de predição genética multivariado (bivariado) em x e y.}$$

Assim, as eficiências relativas são dadas por:

(a) Indireta/Direta

$$\hat{E}_1 = \hat{\psi}_x / \hat{h}_{ay}^2 = -0,2596$$

(b) Bivariada/Direta

$$\hat{E}_2 = \hat{\psi}_{x,y} / \hat{h}_{ay}^2 = \left[1 + \frac{(\hat{h}_{ax} \hat{r}_a / \hat{h}_{ay}) - \hat{r}_f}{1 - \hat{r}_f^2} \right]^{1/2} = 1,1100$$

Verifica-se, no presente caso, que a seleção indireta via DC é totalmente inadequada, devendo conduzir à redução na média de PF.

A seleção bivariada, usando DC como auxiliar, conduz a 11% a mais de ganho genético em PF do que a seleção apenas com base em PF. Mesmo com correlação genética negativa entre os dois caracteres, este resultado pode ser explicado pela grande diferença entre os valores da correlação genética e ambi-

ental (Capítulo 3). A estimativa da correlação ambiental equivale a $\hat{r}_e = \frac{\hat{r}_F - \hat{h}_{ax}\hat{h}_{ay}\hat{r}_a}{\sqrt{[1 - \hat{h}_{ax}^2][1 - \hat{h}_{ay}^2]}} = 0,4126$ e, portanto, $\hat{r}_e - \hat{r}_a = 0,6958 \approx 0,70$.

A eficiência de 1,11 é válida para a seleção massal ou individual. Para outros métodos de seleção mais acurados do que a seleção massal, tal como o índice multi-efeitos e o BLUP, esta eficiência será menor. De fato, para o presente caso, pelo método do índice multiefeitos, a eficiência da seleção bivariada em relação à seleção direta foi de 1,09 (cálculos não apresentados).

Para a seleção individual, os pesos (b_y e b_x) a serem utilizados para os caracteres PF e DC, respectivamente, são dados pelas expressões apresentadas no Capítulo 3. Atribuindo 1 a b_y , b_x é dado por $\hat{b}_x = \frac{\hat{\sigma}_{f_y} \hat{r}_a \hat{h}_{ax} - \hat{r}_f \hat{h}_{ay}}{\hat{\sigma}_{f_x} \hat{h}_{ay} - \hat{r}_a \hat{r}_f \hat{h}_{ax}} = -8,0245$.

Dessa forma, no presente caso, a seleção bivariada atribui peso positivo ao caráter PF e peso negativo ao caráter DC. O peso negativo de DC significa que tal caráter é útil para indicar uma influência ambiental sobre PF. A alta magnitude de b_x em relação a b_y é reflexo da grande diferença de escala entre os dois caracteres, conforme pode ser verificado pelos valores da média geral dos dois caracteres.

A partir dos mesmos dados experimentais, outros parâmetros relevantes podem ser estimados, os quais permitem uma detalhada avaliação ambiental da área experimental. Este assunto é abordado em detalhes no tópico 19.11.

5.1.2. Métodos de seleção

Uma observação em um indivíduo do teste de progénie em questão pode ser descrita pelo seguinte modelo linear: $Y_{ijk} = \mu + b_j + p_i + e_{ij} + d_{ijk}$, em que:

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$;

b_j : efeito de bloco, aleatório, $E(b_j) = 0$ e $E(b_j^2) = \sigma_b^2$;

p_i : efeito da progénie, aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$;

e_{ij} : efeito de parcela ij, aleatório, $E(e_{ij}) = 0$ e $E(e_{ij}^2) = \sigma_e^2$;

d_{ijk} : efeito do indivíduo k dentro da parcela ij, aleatório, $E(d_{ijk}) = 0$ e $E(d_{ijk}^2) = \sigma_d^2$.

Reescrevendo os efeitos do modelo, em termos de desvios, tem-se que:

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

$\bar{Y}_{...}$: efeito da média geral;

$(Y_{ijk} - \bar{Y}_{ij.})$: efeito do indivíduo dentro da parcela;

$(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$: efeito da parcela;

$(\bar{Y}_{i..} - \bar{Y}_{...})$: efeito da progênies;

$(\bar{Y}_{.j.} - \bar{Y}_{...})$: efeito do bloco;

$\bar{Y}_{...}$: média geral do ensaio;

$\bar{Y}_{i..}$: média da família no ensaio;

$\bar{Y}_{ij.}$: média da família em determinado bloco (média da parcela);

Y_{ijk} : valor individual;

$\bar{Y}_{.j.}$: média do bloco.

As covariâncias entre os componentes de cada efeito aleatório do modelo linear com o valor genético aditivo a ser predito são dadas por (Resende, 1991; Resende & Higa, 1994a):

$$COV(a, Y_{ijk}) = 2COV(MF) = \sigma_a^2$$

$$COV(a, \bar{Y}_{ij.}) = \frac{1}{n} 2COV(MF) + \frac{(n-1)}{n} 2COV(MTS \text{ ou } TS) = \frac{1}{n} \sigma_a^2 + \frac{(n-1)\rho_a}{n} \sigma_a^2$$

$$COV(a, \bar{Y}_{i..}) = \frac{1}{nb} 2COV(MF) + \frac{(nb-1)}{nb} 2COV(MTS \text{ ou } TS) = \frac{1}{nb} \sigma_a^2 + \frac{(nb-1)\rho_a}{nb} \sigma_a^2$$

$$COV(a, \bar{Y}_{.j.}) = \frac{1}{np} 2COV(MF) + \frac{(n-1)}{np} 2COV(MTS \text{ ou } TS) = \frac{1}{np} \sigma_a^2 + \frac{(n-1)\rho_a}{np} \sigma_a^2$$

$$COV(a, \bar{Y}_{...}) = \frac{1}{npb} 2COV(MF) + \frac{(nb-1)}{npb} 2COV(MTS \text{ ou } TS) = \frac{1}{npb} \sigma_a^2 + \frac{(nb-1)\rho_a}{npb} \sigma_a^2, \text{ em que:}$$

ρ_a : correlação genética aditiva intraclasso dentro de família: $\rho_a = (1/4)$ para meios irmãos e $\rho_a = (1/2)$ para irmãos germanos (quando são avaliadas famílias aparentadas ρ_a é dado por $\rho_a = (\rho_{ad} - \rho_{ae})/(1 - \rho_{ae})$, em que ρ_{ad} e ρ_{ae} referem-se aos parentescos genéticos aditivos dentro e entre progêneres, respectivamente);

$COV(MF)$: covariância mãe-filho;

$COV(MTS \text{ ou } TS)$: covariância meio tio-sobrinho ou tio-sobrinho, quando as progêneres em avaliação são de meios irmãos ou irmãos germanos, respectivamente.

Efetuando-se as devidas operações, as covariâncias entre os efeitos do modelo linear e o valor genético aditivo são:

(a) Efeito de indivíduo dentro de parcela

$$COV[a, (Y_{ijk} - \bar{Y}_{ij.})] = COV(a, Y_{ijk}) - COV(a, \bar{Y}_{ij.}) = \{[(n-1)(1-\rho_a)]/n\} \sigma_a^2$$

(b) Efeito de progêneres

$$COV[a, (\bar{Y}_{i..} - \bar{Y}_{...})] = COV(a, \bar{Y}_{i..}) - COV(a, \bar{Y}_{...}) = \frac{(p-1)}{p} \frac{[1+(nb-1)\rho_a]}{nb} \sigma_a^2$$

(c) Efeito de parcela

$$\text{COV}[a, (\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_{.j} + \bar{Y}...)] = \text{COV}(a, \bar{Y}_{ij}) - \text{COV}(a, \bar{Y}_i) - \text{COV}(a, \bar{Y}_{.j}) + \text{COV}(a, \bar{Y}...)$$

$$\text{COV}[a, (\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_{.j} + \bar{Y}...)] = \frac{(b-1)}{b} \frac{(p-1)}{p} \frac{1-\rho_a}{n} \sigma_a^2$$

(d) Efeito de bloco

$$\text{COV}[a, (\bar{Y}_{.j} - \bar{Y}...)] = \text{COV}(a, \bar{Y}_{.j}) - \text{COV}(a, \bar{Y}...) = \frac{b-1}{b} \frac{(1-\rho_a)}{np} \sigma_a^2$$

O somatório das 4 covariâncias (itens *a* a *d*) descritas anteriormente equivale a $\sigma_a^2 - \text{COV}(a, \bar{Y}...)$, ou seja, utilizando o índice descrito acima, da variância aditiva total apenas a fração:

$\text{COV}(a, \bar{Y}...) = [1/(npb)] \sigma_a^2 + [(nb-1) \rho_a \sigma_a^2] / (npb) = [(1+(nb-1) \rho_a) \sigma_a^2] / (npb)$ não é explorada na seleção. Isto porque $\bar{Y}...$ é o efeito fixo do modelo e, portanto, constante, de forma que não afeta o mérito relativo dos candidatos à seleção.

Entretanto, pela expressão de $\text{COV}(a, \bar{Y}...)$, constata-se que esta tende a zero, quando os números de indivíduos avaliados nos experimentos atingem magnitudes razoáveis. Em outras palavras, quanto menor for o tamanho da população do experimento, mais variância aditiva se perde na seleção, variância essa que é retida no efeito da média geral. Este é o efeito da amostragem populacional finita, na eficiência da seleção.

De todas as covariâncias descritas nos itens *a* a *d*, aquela referente ao efeito de blocos apresenta o menor valor, já que é função inversa do produto do número de plantas por parcela pelo número de progênies avaliadas. Assim, em experimentos com tamanhos razoáveis, os efeitos de blocos poderão ser negligenciados, visto que a covariância com o valor genético atingirá magnitudes desprezíveis. Essa baixa covariância genética é esperada, visto que a média de blocos não representa nenhuma entidade genética.

Pelo que foi exposto, o efeito de bloco pode ser desconsiderado na seleção e, quando completo, tal efeito pode e deve ser considerado como fixo, quando se utiliza o procedimento BLUP (ver Capítulo 7). Dessa forma, métodos de seleção podem ser estruturados, visando explorar os efeitos de progênies, de parcela e de indivíduos dentro de parcela e, portanto, torna-se necessário conhecer as variâncias destes efeitos, conforme apresentado a seguir:

(a) Efeito de indivíduo dentro de parcela

$$\text{Var}(Y_{ijk} - \bar{Y}_{ij}.) = E(Y_{ijk}^2) - 2E(Y_{ijk}\bar{Y}_{ij}.) + E(\bar{Y}_{ij}^2) = [(n-1)/n]\sigma_d^2$$

(b) Efeito de progênie

$$\text{Var}(\bar{Y}_{i..} - \bar{Y}...) = E(Y_{i..}^2) - 2E(\bar{Y}_{i..}\bar{Y}...) + E(\bar{Y}^2) = [(p-1)/p](\sigma_p^2 + \sigma_e^2/b + \sigma_d^2/nb)$$

(c) Efeito de parcela

$$\text{Var}(\bar{Y}_{ij.} - \bar{Y}_i - \bar{Y}_{.j} + \bar{Y}...) = E(Y_{ij.}^2) + E(\bar{Y}_i^2) + E(\bar{Y}_{.j}^2) + E(\bar{Y}^2) -$$

$$-2E(\bar{Y}_{ij}\bar{Y}_{i..}) - 2E(\bar{Y}_{ij}\bar{Y}_{.j..}) + 2E(\bar{Y}_{ij}\bar{Y}_{...}) + 2E(\bar{Y}_{i..}\bar{Y}_{.j..}) - 2E(\bar{Y}_{i..}\bar{Y}_{...}) - 2E(\bar{Y}_{.j..}\bar{Y}_{...})$$

$$Var(\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j..} + \bar{Y}_{...}) = \frac{(p-1)(b-1)}{p} (\sigma_e^2 + \sigma_d^2/n)$$

O procedimento ótimo (que maximiza a correlação entre os valores genéticos verdadeiros e o preditor destes valores) de seleção é dado pelo índice multiefeitos (Resende & Higa, 1994):

$$\begin{aligned} I &= b_1 d_{ijk} + b_2 p_i + b_3 e_{ij} \\ &= b_1 (\bar{Y}_{ijk} - \bar{Y}_{i..}) + b_2 (\bar{Y}_{i..} - \bar{Y}_{...}) + b_3 (\bar{Y}_{ij..} - \bar{Y}_{i..} + \bar{Y}_{...}) \end{aligned}$$

Os coeficientes b_i são determinados a partir do sistema matricial:

$$\left[\begin{array}{ccc} \frac{(n-1)}{n} \sigma_d^2 & 0 & 0 \\ 0 & \frac{(p-1)}{p} \left(\sigma_p^2 + \frac{\sigma_e^2}{b} + \frac{\sigma_d^2}{nb} \right) & 0 \\ 0 & 0 & \frac{(p-1)(b-1)}{p} \left(\sigma_e^2 + \frac{\sigma_d^2}{n} \right) \end{array} \right] \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \frac{(n-1)(1-\rho_a)}{n} \sigma_a^2 \\ \frac{(p-1)[1+(nb-1)\rho_a]}{nb} \sigma_a^2 \\ \frac{(1-\rho_a)(b-1)(p-1)}{n} \sigma_a^2 \end{bmatrix}$$

Resolvendo-se esse sistema, os coeficientes de ponderação do índice multiefeitos são:

$$b_1 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_d^2} = \text{"herdabilidade" do efeito de indivíduo na parcela};$$

$$b_2 = \frac{1+(nb-1)\rho_a}{nb} \sigma_a^2 = \text{"herdabilidade" do efeito de progênie};$$

$$\frac{\sigma_p^2 + \sigma_e^2 / b + \sigma_d^2 / nb}{\sigma_e^2 + \sigma_d^2 / n}$$

$$b_3 = \frac{[(1-\rho_a)/n] \sigma_a^2}{\sigma_e^2 + \sigma_d^2 / n} = \text{"herdabilidade" do efeito de parcela};$$

Nota-se que o termo "herdabilidade" foi aqui utilizado para denotar a fração do diferencial de seleção retido na descendência.

O índice multiefeitos pode também ser expresso em termos dos ponderadores do valor individual, média da família, média da parcela, média do bloco e média geral, fazendo-se:

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

Por este último índice, nota-se que o efeito de bloco não é considerado na seleção, mas a média do bloco sim, recebendo peso negativo por referir-se, predominantemente, ao efeito ambiental.

Para delineamentos com uma planta por parcela, sem prévia correção dos valores individuais para os efeitos de blocos, o índice é dado por:

$I = b_2^*(\bar{Y}_{i..} - \bar{Y}_{...}) + b_3^*(Y_{ij} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...})$ em que b_2^* e b_3^* são obtidos fazendo-se $n=1$ em b_2 e b_3 respectivamente.

Para o caso balanceado, o método do índice multiefeitos equivale ao procedimento BLUP individual, conforme demonstrado por Resende & Fernandes, (1999).

No experimento com progênies de meios-irmãos de acerola, os seguintes procedimentos de seleção direta podem ser comparados, visando ao melhoramento do caráter produção de frutos, em termos dos coeficientes (b_i) de ponderação das informações fenotípicas e acurácia seletivas (r_{la} ou r_{aa}).

(i) Seleção Massal

Coeficiente de ponderação do valor individual

$$\hat{b}_i = \hat{h}_{ay}^2 = 0,3147$$

Acurácia

$$\hat{r}_{aa} = \hat{h}_{ay} = 0,5610$$

(ii) Seleção pelo índice multiefeitos

Coeficiente de ponderação:

$$\hat{b}_1 = \frac{(3/4) \hat{\sigma}_a^2}{\hat{\sigma}_d^2} = \frac{(3/4) \hat{h}_a^2}{1 - [(1/4) \hat{h}_a^2 + \hat{c}^2]} = 0,2821$$

$$\hat{b}_2 = \frac{[(3+nb)/(4nb)] \hat{\sigma}_a^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 / b + \hat{\sigma}_d^2 / nb} = \frac{[1 + (nb-1)(1/4)] \hat{h}_a^2}{1 + (n-1)[(1/4) \hat{h}_a^2 + \hat{c}^2] + (b-1)n(1/4) \hat{h}_a^2} = 0,5568$$

$$\hat{b}_3 = \frac{[3/(4n)] \hat{\sigma}_a^2}{\hat{\sigma}_d^2 / n + \hat{\sigma}_e^2} = \frac{(3/4) \hat{h}_a^2}{1 + (n-1)[(1/4) \hat{h}_a^2 + \hat{c}^2] - n(1/4) \hat{h}_a^2} = 0,2008$$

O índice multiefeitos, no presente caso, é dado então por

$$\hat{I} = \hat{b}_1 Y_{ijk} + (\hat{b}_2 - \hat{b}_3) \bar{Y}_{i..} + (\hat{b}_3 - \hat{b}_1) \bar{Y}_{ij..} - b_3 \bar{Y}_{..j} + (\hat{b}_3 - \hat{b}_2) \bar{Y}_{...} =$$

$$= 0,2821 Y_{ijk} + 0,3560 \bar{Y}_{i..} - 0,0813 \bar{Y}_{ij..} - 0,2008 \bar{Y}_{..j} - 0,3560 \bar{Y}_{...}$$

Verifica-se que a média do bloco, a média geral e a média da parcela receberam pesos negativos (significando que tais valores fenotípicos refletem mais os efeitos ambientais do que genéticos), ao passo que o valor individual e a média da progénie receberam pesos positivos, pois informam mais sobre o valor genético.

Acurácia

$$\hat{r}_{aa} = [\alpha_1 \hat{b}_1 + \alpha_2 \hat{b}_2 + \alpha_3 \hat{b}_3]^{1/2} = 0,5955$$

$$\text{pois: } \alpha_1 = (1 - \rho_a) \frac{n-1}{n} = 0,75 (3/4) = 0,5625$$

$$\alpha_2 = \frac{1 + (nb - 1)\rho_a}{nb} \frac{p - 1}{p} = 0,3075$$

$$\alpha_3 = \frac{(1 - \rho_a)}{n} \frac{p - 1}{p} \frac{b - 1}{b} = \frac{0,75}{4} \frac{61}{62} \frac{2}{3} = 0,1230$$

(iii) Seleção pelo efeito de progêneres (seleção entre progêneres sem sementes remanescentes)

Coeficiente de ponderação da média de família

$$\hat{b}_2 = 0,5568$$

Acurácia

$$\hat{r}_{aa} = [\alpha_2 \hat{b}_2]^{1/2} = 0,4138$$

(iv) Seleção pelo efeito de indivíduo na parcela (seleção dentro de progêneres)

Coeficiente de ponderação

$$\hat{b}_1 = 0,2821$$

Acurácia

$$\hat{r}_{aa} = [\alpha_1 \hat{b}_1]^{1/2} = 0,3983$$

(v) Seleção de Irmãos (seleção entre progêneres com sementes remanescentes)

Coeficiente de ponderação da média de família

$$\hat{b}_4 = \frac{(1/4) \hat{\sigma}_a^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 / b + \hat{\sigma}_d^2 / nb} = \frac{(1/4) nb \hat{h}_a^2}{1 + (n-1)[(1/4) \hat{h}_a^2 + \hat{c}_2] + (b-1) n (1/4) \hat{h}_a^2} = 0,4454$$

Acurácia

$$\hat{r}_{aa} = \left[\rho_a \hat{b}_4 \frac{p - 1}{p} \right]^{1/2} = 0,3310$$

(vi) Seleção e recombinação de genitores com base no teste de suas progêneres (seleção de genitores)

Coeficiente de ponderação da média de progênie

$$\hat{b}_5 = \frac{(1/2) \hat{\sigma}_a^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 / b + \hat{\sigma}_d^2 / nb} = 2\hat{b}_4 = 0,8908$$

Acurácia

$$\hat{r}_{aa} = \left[\left(1/2 \right) \hat{b}_5 \frac{p-1}{p} \right]^{1/2} = 0,6620$$

(vii) Seleção em dois estágios: entre e dentro de progênies.

Seleção entre progênies

$$\hat{b}_2 = 0,5568$$

$$\hat{r}_{aa} = 0,4138$$

Seleção dentro de progênies

$$\hat{b}_1 = 0,2821$$

$$\hat{r}_{aa} = 0,3983$$

(viii) Seleção por um índice imperfeito baseado em p_i e d_{ijk} (Bueno Filho, 1992; Resende & Higa, 1994b; Pires et al., 1996).

Índice

$$I = \hat{b}_1 (Y_{ijk} - \bar{Y}_{ij\cdot}) + \hat{b}_2 (\bar{Y}_{i\cdot} - \bar{Y}\dots)$$

Coeficientes de Ponderação

$$\hat{b}_1 = 0,2821; \hat{b}_2 = 0,5568$$

Acurácia

$$\hat{r}_{aa} = [\alpha_1 \hat{b}_1 + \alpha_2 \hat{b}_2]^{1/2} = 0,5744$$

(ix) Seleção por um índice imperfeito modificado (Pires et al., 1996).

Índice

$$I = \hat{b}_6 (Y_{ijk} - \bar{Y}_{\cdot j\cdot}) + \hat{b}_7 (\bar{Y}_{i\cdot\cdot} - \bar{Y}\dots)$$

Coeficientes de ponderação

Neste caso, os coeficientes \hat{b}_6 e \hat{b}_7 são dados por:

$$\begin{bmatrix} \hat{b}_6 \\ \hat{b}_7 \end{bmatrix} = \begin{bmatrix} \sigma_{C_1}^2 & \sigma_{C_1 C_2} \\ \sigma_{C_1 C_2} & \sigma_{C_2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{C_1 a} \\ \sigma_{C_2 a} \end{bmatrix}, \text{ em que:}$$

$$\sigma_{c_1}^2 = Var(Y_{ijk} - \bar{Y}_{\cdot j\cdot}) = \frac{p-1}{p} (\sigma_p^2 + \sigma_e^2) + \frac{np-1}{np} \sigma_d^2 = 126,2833$$

$$\sigma_{c_2}^2 = Var(\bar{Y}_{i\cdot\cdot} - \bar{Y}\dots) = \frac{p-1}{p} (\sigma_p^2 + \sigma_e^2 / b + \sigma_d^2 / nb) = 22,0749$$

$$\sigma_{c_1 c_2} = \sigma_{c_1 c_2} = \sigma_{c_2}^2 = \frac{p-1}{p} \left(\sigma_p^2 + \sigma_e^2 / b + \sigma_d^2 / nb \right) = 22,0749$$

$$\sigma_{c_1 a} = \frac{(np-1)-(n-1)}{np} \rho_a \sigma_a^2 = 39,6947$$

$$\sigma_{c_2 a} = \frac{p-1}{p} \frac{1+(nb-1)}{nb} \rho_a \sigma_a^2 = 12,2912$$

Resolvendo, tem-se $\hat{b}_6 = 0,2630$ e $\hat{b}_7 = 0,2938$

Acurácia

$$r_{aa} = [\alpha_6 \hat{b}_6 + \alpha_7 \hat{b}_7]^{1/2} = 0,5928, \text{ pois}$$

$$\alpha_6 = \frac{(np-1)-(n-1)}{np} \rho_a = 0,9929$$

$$\alpha_7 = \frac{p-1}{p} \frac{1+(nb-1)}{nb} \rho_a = 0,3075$$

Os métodos de seleção de indivíduos que propiciam a mesma intensidade de seleção podem ser comparados, conforme apresentado a seguir:

Método	Acurácia	Eficiência sobre a seleção massal	Eficiência do IME sobre os demais métodos
(ii) Índice multiefeitos (IME)	0,5955	1,06	1,00
(ix) Índice imperfeito modificado	0,5928	1,05	1,005
(viii) Índice imperfeito	0,5743	1,02	1,04
(i) Seleção massal	0,5610	1,00	1,06

Os resultados comprovam a superioridade teórica do índice multiefeitos (IME) como procedimento ótimo. Os demais índices podem, em algumas situações, gerar resultados próximos aos obtidos pelo índice multiefeitos, mas serão, na melhor das hipóteses, iguais ao índice multiefeitos. Ademais, o índice multiefeitos é o próprio BLUP para a situação de dados balanceados, fato que não o são os demais índices imperfeitos (Resende, 1999b; Resende & Fernandes, 1999). Dessa forma, a utilização de tais índices não é justificada.

A eficiência do IME em relação à seleção massal não foi tão elevada no presente exemplo, devido a dois fatores: (i) herdabilidade relativamente alta do caráter; (ii) pequeno número de indivíduos por família. Essa eficiência pode ser aumentada elevando-se o número de repetições no experimento. Por exemplo, se fossem utilizadas 10 repetições (40 indivíduos por família, em vez de 12), a

eficiência passaria de 1,06 para cerca de 1,12 (Resende, 1995), fato que pode ser bastante compensador.

As comparações dos métodos de seleção, que propiciam diferentes intensidades de seleção, devem basear-se no ganho genético. Considerando a seleção dos 30 melhores indivíduos pelo método IME, dos 3 melhores indivíduos de cada uma das 10 melhores famílias pela seleção entre e dentro de famílias, 62 (o melhor de cada família) indivíduos pela seleção dentro de famílias e de 10 genitores, 10 famílias e 10 grupos de irmãos pela seleção entre genitores, entre famílias e seleção de irmãos, respectivamente, obtiveram-se os seguintes ganhos genéticos, empregando-se o software Selegen (Resende et al., 1994a):

Método	Acurácia	Ganho genético (kg)	Ganho genético (%)
(ii) Índice multiefeitos	0,5955	9,47	44,10
(vii) Entre e dentro de família	(0,4138; 0,3983)	7,89	36,73
(iv) Dentro de família	0,3983	4,35	20,25
(iii) Entre famílias	0,4138	4,37	20,34
(v) Seleção de irmãos	0,3310	3,49	16,25
(vi) Seleção de genitores	0,6620	6,99	32,54

Verifica-se que a seleção e recombinação dos genitores, embora seja o método mais acurado, conduziu a menor ganho genético que o IME, pois propiciou menor intensidade de seleção. De todos os métodos, a seleção de irmãos usando sementes remanescentes propiciou o menor ganho. O IME propiciou o maior ganho, resultando em 20% (9,47/7,89) a mais de ganho do que a seleção entre e dentro de progêniens.

O ganho genético equivale à média dos valores genéticos preditos (expressos como desvios da média geral) dos indivíduos selecionados. Assim, no caso da seleção de genitores, têm-se:

Ordens dos melhores genitores	Nº dos genitores	Valor genético predito (kg)	Ganho genético acumulado (kg)	Ganho acumulado (%)
1	51	13,80	13,80	64,24
2	59	11,78	12,79	59,54
3	20	8,40	11,33	52,74
4	92	7,53	10,38	48,32
5	63	5,63	9,43	43,90
6	66	5,46	8,77	40,83
7	91	4,84	8,21	38,22
8	56	4,28	7,72	35,94
9	12	4,18	7,32	34,08
10	47	3,97	6,99	32,54

Uma maneira de aumentar a eficiência da seleção de genitores, sem comprometer a recombinação nos pomares e o tamanho efetivo, consiste em adotar a prática da colheita de sementes das matrizes com os maiores valores genéticos preditos. No exemplo da acerola, com base no teste de progênies das 62 matrizes, as 10 melhores poderiam ser clonadas em um pomar de sementes, mas a colheita das sementes para plantio poderia ser realizada somente nas melhores destas 10. Assim, ter-se-ia que:

(i) Colheita de sementes da melhor matriz (plantio monoprogénie)

$$\text{Ganho genético (kg)} = \frac{1}{2}(13,80 + 6,99) = 10,40 \text{ kg}$$

$$\text{Ganho genético (\%)} = (10,40 / 21,48) \times 100 = 48,42\%$$

(ii) Colheita de sementes das 5 melhores matrizes

$$\text{Ganho genético (kg)} =$$

$$\frac{1}{2} \left\{ [(13,80 + 11,78 + 8,40 + 7,53 + 5,63) / 5] + 6,99 \right\}$$

$$= \frac{1}{2}(9,43 + 6,99) = 8,21 \text{ kg}$$

$$\text{Ganho genético (\%)} = (8,21 / 21,48) \times 100 = 38,22\%$$

Verifica-se que esta estratégia é muito eficiente para aumentar o ganho genético. No caso, o plantio monoprogénie conduzirá a 10% (48,42/44,10) a mais de ganho do que o IME. A implementação prática desta estratégia requer a clonagem da melhor matriz em um grande número de rametes, de forma que se possa produzir a quantidade de sementes necessária.

5.1.3. Tamanho efetivo populacional e número efetivo de famílias selecionadas

Sendo N_f o número de famílias selecionadas e k_f o número de indivíduos selecionados por família, e considerando as 62 matrizes como não aparentadas, as seguintes estimativas para o tamanho efetivo populacional de variância ($N_{e(v)}$) são encontradas.

(i) Seleção entre famílias

Tamanho efetivo de cada família

$$N_e = \frac{4k_f}{k_f + 3} = \frac{4 \times 12}{12 + 3} = 3,2$$

Tamanho efetivo associado à seleção das 10 melhores famílias

$$N_e = \frac{4N_f k_f}{k_f + 3} = 10 \times 3,2 = 32$$

(ii) Seleção de genitores

$N_e = N = 10$, onde N é o número de genitores selecionados.

(iii) Seleção dentro de famílias

Com a seleção de 1 indivíduo por família, $N_e = N_f = 62$.

(iv) Seleção entre e dentro de famílias

Com $k_f = 3$ indivíduos selecionados em cada uma das 10 melhores famílias:

$$N_e = \frac{4 \cdot 10 \cdot 3}{6} = 20$$

(v) Seleção pelo índice multiefeitos ou outro índice ou seleção massal

$$N_e = \frac{4N_f \bar{k}_f}{\bar{k}_f + 3 + (\sigma_{k_f}^2 / \bar{k}_f)}, \text{ em que:}$$

\bar{k}_f = número médio de indivíduos selecionados por família.

$\sigma_{k_f}^2$ = variância do número de indivíduos selecionados por família.

Para ilustrar o cálculo, serão considerados os 10 melhores indivíduos selecionados pelo índice multiefeitos:

Ordem	Bloco	Família	Indivíduo/ parcela	VG(Kg)	Ganho acumulado (Kg)	Ganho Ac (%)	Média da população melhorada	N_e
1	1	20	4	13,05	13,05	60,75	34,53	1,00
2	2	63	4	12,71	12,88	59,96	34,36	2,00
3	1	51	4	11,70	12,49	58,14	33,97	3,00
4	1	92	4	11,10	12,14	56,51	33,62	4,00
5	1	59	1	11,03	11,92	55,49	33,40	5,00
6	2	51	3	10,94	11,75	54,70	33,24	5,50
7	3	59	1	10,87	11,63	54,14	33,11	6,07
8	1	51	3	10,69	11,51	53,58	32,99	6,27
9	1	56	3	10,69	11,42	53,16	32,90	7,25
10	1	51	1	10,67	11,34	52,79	32,83	7,21

Para o cômputo do N_e associado à seleção dos 10 melhores indivíduos, obtém-se a seguinte contribuição de indivíduos por cada família:

Família	K_f
20	1
63	1
51	4
92	1
59	2
56	1
Total	$N = 10$

Assim:

$$\bar{k}_f = N/N_f = 10/6 = 1,6667$$

$$\sigma_{k_f}^2 = \left[\sum k_f^2 - \frac{(\sum k_f)^2}{N_f} \right] / (N_f - 1) = [24 - 10^2/6]/5 = 1,4667$$

Dessa forma:

$$N_e = \frac{4.6.1,6667}{1,6667 + 3 + (1,4667/1,6667)} = 7,2116$$

O número efetivo de famílias selecionadas é dado por:

$$N_{ef} = (\sum k_f)^2 / \sum k_f^2 = 10^2 / 24 = 4,1667$$

Assim, o número físico de famílias selecionadas foi 6, mas o número efetivo de famílias foi de apenas 4,17, devido à variação no número de indivíduos contribuídos por família. Na seleção de igual número de indivíduos por família $N_f = N_{ef}$

A diversidade genética (D) após a seleção pode ser quantificada conforme Wei & Lindgren (1996):

$$D = \frac{N_{ef}}{N_{fo}} = \frac{4,16}{62} = 0,0671, \text{ em que } N_{fo} \text{ refere-se ao número original de famílias e } 0 < D \leq 1.$$

Desse modo, D próximo de zero indica a quase extinção da variabilidade genética contida na população formada por N_{fo} famílias e D próximo de 1 indica a quase manutenção da variabilidade total da população de referência.

5.1.4. Tamanho efetivo em função da proporção de utilização dos indivíduos

Os indivíduos clonados em um pomar de sementes podem ser repetidos em números diferentes de ramegos de cada clone. Em função disto, os tamanhos efetivos computados anteriormente não são válidos.

Tomando por base a seleção de genitores não aparentados a serem re-combinados em um pomar de sementes, mas representados por diferentes números (m) de ramegos de cada um, o tamanho efetivo é dado por:

$$N_e = (\sum m_i)^2 / \sum m_i^2 = 1 / \sum f_i^2, \text{ em que } f_i \text{ é a freqüência ou proporção do genitor } i, \text{ dada por } f_i = m_i / \sum m_i.$$

Dessa forma, se dos 10 genitores, o melhor for repetido 8 vezes no pomar e os demais forem repetidos 4 vezes, colhendo-se sementes (em quantidades aproximadamente iguais) de todos os ramegos, têm-se:

$$f_1 = 8/44 = 0,1818$$

$$f_2 \text{ a } f_{10} = 4/44 = 0,0909$$

$$N_e = 1 / \sum f_i^2 = 1 / (0,1818^2 + 9 \times 0,0909^2) = 9,31, \text{ ou}$$

$$N_e = (\sum m_i)^2 / \sum m_i^2 = 44^2 / (64 + 9 \times 16) = 9,31$$

Neste caso, o tamanho efetivo reduz-se de 10 para 9,31.

Quando os indivíduos no pomar são aparentados (advindos da seleção pelo IME ou outro método de seleção individual) e são repetidos em diferentes números de ramos no pomar, o tamanho efetivo corrigido pela proporção de utilização dos indivíduos pode ser dado pelas expressões apresentadas por Resende & Resende (1997).

Considerando a seleção dos 10 indivíduos pelo IME (resultados apresentados anteriormente) e utilizando-os com os números r_i de ramos 10,9,8,..., 2 e 1, têm-se:

Indivíduo	Família	Nº de ramos (ri)	Proporção de utilização	Utilização acumulada	Ganho genético
1	20	10	0,1818	0,1818	-
2	63	9	0,1636	0,3454	-
3	51	8	0,1455	0,4909	-
4	92	7	0,1273	0,6182	-
5	59	6	0,1091	0,7273	-
6	51	5	0,0909	0,8182	-
7	59	4	0,0727	0,8909	-
8	51	3	0,0545	0,9454	-
9	56	2	0,0364	0,9818	-
10	51	1	0,0182	1,0000	11,70 (54,47%)

A contribuição de cada família, em termos de número físico de indivíduos (k_f), número de ramos (r_f) e número efetivo de indivíduos corrigido para a proporção de utilização (N_{pi}), pode ser assim representada:

Família	K _f	r _f	N _{pi}
20	1	10	1
63	1	9	1
51	4	17	2,9192
92	1	7	1
59	2	10	1,9231
56	1	2	1
N = 10		$\Sigma r_f = 55$	N_p = 8,8423

O N_{pi} de cada família foi assim obtido:

$$N_{pi} \text{ da família } 51 = (\sum r_i)^2 / \sum r_i^2 = (8+5+3+1)^2 / (8^2 + 5^2 + 3^2 + 1^2) = 2,9192$$

$$N_{pi} \text{ da família } 59 = (6+4)^2 / (6^2 + 4^2) = 1,9231$$

O número efetivo de famílias corrigido (N_{efc}) para a proporção de utilização dos indivíduos equívale a:

$$N_{efc} = (\sum r_f)^2 / \sum r_f^2 = 55^2 / 623 = 4,8555$$

\bar{k}_f corrigido (\bar{k}_{fc}) e σ_{kfc}^2 corrigido (σ_{kfc}^2) equívalem a :

$$\bar{k}_{fc} = N_p / N_f = 8,8423 / 6 = 1,4737$$

$$\sigma_{kfc}^2 = \left[\sum N_{pi}^2 - \frac{(\sum N_{pi})^2}{N_f} \right] / (N_f - 1) = \left[16,22 - \frac{8,8423^2}{6} \right] / 5 = 0,6378$$

Dessa forma, o N_e corrigido equívale

$$N_{ec1} = \frac{4N_f \bar{k}_{fc}}{\bar{k}_{fc} + 3 + (\sigma_{kfc}^2 / \bar{k}_{fc})} = \frac{4 \cdot 6 \cdot 1,4737}{1,4737 + 3 + (0,6378 / 1,4737)} = 7,2086$$

Verifica-se que a utilização dos indivíduos selecionados em diferentes proporções manteve o mesmo tamanho efetivo ($\approx 7,0$) e conduziu a 3% ($54,47/52,79$) a mais de ganho genético do que a recombinação balanceada. Isto prova que a recombinação desbalanceada pode ser uma estratégia efetiva de aumento da taxa de melhoramento genético.

Embora a recombinação desbalanceada tenha conduzido a um menor número médio de indivíduos selecionados por família (1,4737 contra 1,6667), a mesma conduziu a um maior número efetivo de famílias (4,8555 contra 4,1667) e menor variância do número de indivíduos selecionados por família (0,6378 contra 1,4667) do que a recombinação balanceada. Estes resultados conduziram a tamanhos efetivos idênticos pelas duas estratégias. É importante ressaltar que o número efetivo de família é correlacionado negativamente com a variância σ_{kf}^2 , sendo que o cômputo de N_{ef} já leva em consideração a variância σ_{kf}^2 .

Outras fórmulas aproximadas mais simples para o cômputo do N_e corrigido são:

$$(i) N_{ec2} = N_e N_p / N$$

$$= 7,2116 \cdot 8,8423 / 10 = 6,38$$

$$(ii) N_{ec3} = \frac{4N_{efc} N_p / N_{efc}}{N_p / N_{efc} + 3} = \frac{4N_{efc} N_p}{3N_{efc} + N_p} = \frac{4 \times 4,8555 \times 8,8423}{3 \times 4,8555 + 8,8423} = 7,34$$

5.1.5. Endogamia da população melhorada e correção do ganho genético

O conhecimento do tamanho efetivo de endogamia ($N_{e(l)}$) permite estimar o coeficiente de endogamia da população recombinada (geração seguinte) pela relação $F=1/(2N_e)$. A prova desta relação pode ser verificada neste exemplo.

Na presente situação, sabe-se que a endogamia deverá advir de autofecundação dos indivíduos e do cruzamento entre os meios-irmãos dentro de cada família. Considerando n_i o número de indivíduos selecionados em cada uma de p famílias, a configuração dos cruzamentos (incluindo os recíprocos) envolvendo os indivíduos selecionados é a seguinte:

- Número total de cruzamentos:

$$\left(\sum_{i=1}^p n_i \right)^2 = \text{arranjos com reposição dos } \sum_{i=1}^p n_i \text{ indivíduos dois a dois;}$$

- Número total de cruzamentos entre meios-irmãos:

$$\sum_{i=1}^p n_i(n_i - 1) = \text{somatório dos arranjos sem reposição dos } n \text{ indivíduos dentro de famílias, tomados 2 a 2.}$$

- Número de autofecundações: $\sum_{i=1}^p n_i$
- Coeficiente de endogamia de descendentes de cruzamento de meios-irmãos: $F_M = (1/8)$
- Coeficiente de endogamia de indivíduos obtidos por autofecundação $F_I = (1/2)$

A endogamia total (F) equivale a:

$$F = \frac{\sum_{i=1}^p n_i(n_i - 1)}{\left(\sum_{i=1}^p n_i \right)^2} F_M + \frac{\sum_{i=1}^p n_i}{\left(\sum_{i=1}^p n_i \right)^2} F_I$$

Considerando a seleção entre e dentro de progênie com $p=10$ famílias selecionadas e $n_i=3$ indivíduos por progênie, tem-se:

$$\begin{aligned} F &= 10 \frac{[3(3-1)]}{30^2} (1/8) + \frac{30}{30^2} (1/2) \\ &= 0,06667(1/8) + 0,0333(1/2) \\ &= 0,025 \end{aligned}$$

O N_e associado é $N_e = \frac{4N_f k_f}{k_f + 3} = 20$ e o $F = 1/(2N_e) = 1/40 = 0,025$, portanto idêntico ao obtido pelo cálculo direto de F (neste caso $N_{e(v)} = N_{e(I)}$).

O N_e de endogamia pode também ser obtido a partir do F , em caso de contribuição desbalanceada por família. No caso da seleção dos 10 melhores indivíduos pelo IME, tem-se.

$$\begin{aligned} F &= \frac{\sum_{i=1}^p n_i(n_i - 1)}{\left(\sum_{i=1}^p n_i\right)^2} F_M + \frac{\sum_{i=1}^p n_i}{\left(\sum_{i=1}^p n_i\right)^2} F_I \\ &= \left[\frac{4(4-1)}{10^2} + \frac{2(2-1)}{10^2} \right] (1/8) + \frac{10}{10^2} (1/2) \\ &= 0,14(1/8) + 0,10(1/2) \\ &= 0,0675 \end{aligned}$$

O N_e de endogamia é então $1/(2 \cdot 0,0675) = 7,4074$, o qual difere ligeiramente do N_e de variância (7,2115).

Outra questão que deve ser mencionada é que a expressão $F=1/(2N_e)$ é válida somente para cruzamentos aleatórios entre os indivíduos recombinados. Com cruzamentos preferenciais ou desbalanceados, a igualdade não se verifica. Considere, por exemplo, que dois indivíduos não aparentados de uma população monóica sejam incluídos em um pomar de sementes (pomar biclonal). O tamanho efetivo equivalente aos dois indivíduos é, logicamente, 2. Incluindo-os em proporção iguais de ramos, tem-se que a probabilidade de autofecundação é $1/N = 1/2$ e a endogamia é, portanto, $F=(1/N)(F_I) = (1/2)(1/2) = 1/4$, quantidade que equivale exatamente a $1/(2N_e)$.

Por outro lado, incluindo um deles na freqüência de 1/11 e o outro na freqüência de 10/11, visando colher sementes apenas da fração 1/11, tem-se que o coeficiente de endogamia dos indivíduos da geração de plantio equivale a $F=(1/N)(F_I) = (1/11)(1/2) = 1/22$ e não a 1/4.

A endogamia da população melhorada deve ser levada em conta no cômputo do ganho genético. O ganho genético corrigido (G_{SC}) para a endogamia é dado por Resende & Bertolucci (1995):

$$G_{SC} = G_S - FH, \text{ em que:}$$

G_S : ganho genético não corrigido

F : coeficiente de endogamia

$H = M_{PFO} - M_{PF1}$: depressão endogâmica

M_{PFO} : média da população melhorada com $F = 0$, expressa como desvio da média geral ($M_{PFO} = G_S$)

M_{PF_1} : média da população melhorada com $F = 1$, expressa como desvio de média geral.

Considerando a depressão endogâmica na mesma taxa de F (ou seja, se $F=0,10$, $H=10\%$ e, assim, sucessivamente), $M_{PF_1} = 0$ e, portanto:

$$\begin{aligned} G_{SC} &= G_S - F(M_{PFO} - M_{PF_1}) \\ &= G_S - FM_{PFO} \\ &= G_S - FG_S \\ &= (1 - F)G_S \\ &= [1 - 1/(2N_e)]G_S \end{aligned}$$

No caso dos 10 indivíduos selecionados pelo IME:

Ordem	Ganho genético (Kg)	N_e	F	Ganho corrigido (kg)
1	13,05	1,00	0,50	6,52
2	12,88	2,00	0,25	9,66
3	12,49	3,00	0,17	10,41
4	12,14	4,00	0,125	10,62
5	11,92	5,00	0,10	10,73
6	11,76	5,50	0,09	10,69
7	11,63	6,07	0,08	10,67
8	11,51	6,27	0,08	10,59
9	11,42	7,25	0,07	10,63
10	11,35	7,21	0,07	10,56

Verifica-se que o ganho genético corrigido é maximizado (10,73) com a seleção dos 5 melhores indivíduos, mesmo com N_e de apenas 5 e $F=0,10$. Por enquanto, seria indicado o estabelecimento de um pomar de sementes com estes 5 indivíduos. Outra opção seria o uso do pomar biclonal desbalanceado formando pelos 2 primeiros indivíduos. Neste caso ter-se-ia: $F = (1/11) (1/2) = 1/22 = 0,045$ e $G_{SC} = 12,88 \cdot 0,955 = 12,30$.

Uma outra necessidade de correção do ganho genético estimado refere-se a perdas devido à interação genótipo x ambiente dentro de um mesmo local. Detalhes a respeito desta correção são apresentados no capítulo 19.

5.1.6. Variância do ganho genético e número de indivíduos a serem selecionados

A determinação do número adequado de indivíduos a serem selecionados deve considerar não só a endogamia da população melhorada mas também a variância do ganho genético, a qual indica a confiabilidade do ganho predito.

A variância do ganho genético ou resposta à seleção (R) pode ser dada por (Resende et al. 1995b):

$\hat{V}(G_s) = \frac{1}{N} \hat{\sigma}_a^2 (1 - \hat{r}_{la}^2)$ ou $\hat{V}(\hat{R}) = \hat{\sigma}_a^2 / N_e$, assumindo os componentes de variância como conhecidos ou estimados com precisão.

Pelas expressões apresentadas, verifica-se que a variância do ganho genético reduz-se à medida que aumenta o número de indivíduos selecionados. Por outro lado, o próprio ganho estimado reduz-se na medida do aumento do número de indivíduos selecionados, em virtude da redução na intensidade de seleção. Assim, deve-se encontrar um número ótimo de indivíduos que maximize o limite inferior do intervalo de confiança (LIIC) do ganho genético corrigido para a endogamia.

A seguir são apresentados os intervalos de confiança (IC) associados aos ganhos genéticos corrigidos, envolvendo a seleção dos 15 melhores indivíduos pelo IME ($r_{aa} = 0,5955$) e dados por $G_{sc} \pm t[V(G_s)]^{1/2}$, em que $t = 1,96$ para 95% de confiança.

Ordem	Ganho corrigido (kg)	L. Inferior I.C. (kg)	L. Superior I.C. (kg)
1	6,53	-3,43	16,49
2	9,66	2,62	16,70
3	10,37	4,62	16,12
4	10,62	5,64	15,60
5	10,73	6,28	15,18
6	10,70	6,64	14,76
7	10,70	6,94	14,46
8	10,59	7,07	14,11
9	10,62	7,30	13,94
10	10,56	7,41	13,71
11	10,53	7,53	13,53
12	10,44	7,57	13,31
13	10,30	7,54	13,06
14	10,17	7,51	12,83
15	10,11	7,54	12,68

Verifica-se que a maximização do limite inferior do intervalo de confiança do ganho genético corrigido para a endogamia ocorre com a seleção dos 12 melhores indivíduos. Dessa forma, 12 seria o número de indivíduos recomendado para o estabelecimento do pomar de sementes. O intervalo de confiança do ganho associado ao pomar bicolonial (ganho corrigido de 12,30) equivale a 5,26 – 19,34, de forma que o limite inferior encontra-se aquém do limite inferior (7,57) do ganho com a recombinação dos 12 melhores. Com a formação deste pomar, pode-se afirmar com 95% de confiança, que o verdadeiro ganho genético será igual ou superior a 35,24% (7,57/21,4815). É importante ressaltar que este ganho pode ser aumentado em cerca de 3%, formando-se o pomar com contribui-

ção desbalanceada dos indivíduos em termos de número de ramações, conforme demonstrado no tópico 5.1.4.

No pomar formado pelos genitores, a maximização do limite inferior do intervalo de confiança do ganho genético corrigido para a endogamia ocorreu com a seleção dos 4 melhores genitores (intervalo: 4,44-13,73kg). Isto ocorreu porque os dois melhores genitores são muito superiores aos demais. Dessa forma, recomenda-se a formação de um pomar com sobreposição de gerações, incluindo os 10 melhores indivíduos selecionados pelo IME mais os dois genitores superiores. Esta estratégia conduz a $G_S = 11,59\text{kg}$; $G_{SC} = 10,54$; $F = 0,0903$; $N_e = 5,54$ e L.I.I.C. = 7,68 (35,75%).

5.1.7. Restrição ao tamanho efetivo e melhoramento no longo prazo

Para o melhoramento da população no longo prazo, deve-se manter um tamanho efetivo compatível com a obtenção do limite seletivo da população. Este tamanho efetivo compatível situa-se ao redor de 50 (ver capítulo 20). Dessa forma, a seleção deve ser restrita em função do N_e de 50. Algumas opções para atingir tal objetivo são: (i) redução da intensidade de seleção; (ii) restrição no número (n) de indivíduos selecionados por família; (iii) restrição na variância σ_n^2 do número de indivíduos selecionados por família. Em geral, a própria restrição em n conduz a redução em σ_n^2 também.

No presente exemplo, empregando-se o software Selegen, obtiveram-se os seguintes resultados.

Método	N	N_e	$G_S (\text{kg})$	$G_S (\%)$	n max	LIIC.
Seleção massal	73	50	6,72	31,28	Irrestrito	5,45 (25,37%)
IME	103	50	6,54	30,44	Irrestrito	5,49 (25,56%)
IME	68	50	6,75	31,42	3	5,48 (25,51%)
IME	76	50	6,76	31,47	4	5,55 (25,84%)

Constata-se que as várias estratégias conduziram a resultados similares, em termos do limite inferior do intervalo de confiança (L.I.I.C.) do ganho genético corrigido para a endogamia. Cabe ao melhorista optar pela melhor alternativa. A recombinação dos 103 melhores indivíduos pelo IME e geração de 103 novas famílias de meios-irmãos, para avaliação no próximo ciclo seletivo, deverá conduzir a um ganho genético de pelo menos 5,49kg ou 25,56%, ao mesmo tempo em que garantirá variabilidade genética suficiente para o melhoramento ao longo prazo. A utilização das duas populações selecionadas (população de produção de sementes formada pelos 12 melhores indivíduos selecionados pelo IME e população de melhoramento formada pelos 103 melhores indivíduos pelo IME) mostrou ser uma estratégia adequada para maximizar os ganhos genéticos no curto prazo (7,57kg ou 35,24%) e no longo prazo (5,49kg ou 25,56%). A

superioridade relativa dos ganhos (35,24% contra 25,56%) justifica o estabelecimento das duas populações com finalidades distintas.

A utilização dos 76 melhores indivíduos pelo IME (com restrição) conduz praticamente ao mesmo ganho que a utilização dos 103 melhores (sem restrição). Entretanto, limita o número de famílias a ser avaliado e a intensidade de seleção entre famílias no ciclo de seleção subsequente.

5.1.8. Variância das estimativas dos parâmetros genéticos

Estimativas precisas de parâmetros genéticos são imprescindíveis aos procedimentos de predição de valores genéticos. Quanto à precisão necessária, tem sido observado que os índices multiefeitos univariados são robustos a erros na estimativa da verdadeira herdabilidade. Os índices baseados em vários caracteres são muito menos robustos, conduzindo a piores resultados do que a seleção baseada em um só caráter, quando as estimativas de correlação genética não são precisas (Sales & Hill, 1976 a e b).

Os estimadores aproximados para a variância das estimativas de herdabilidade e correlação genética são (Falconer, 1989):

$$\hat{V}(\hat{h}_a^2) = 32\hat{h}_a^2 / (nbp);$$

$$\hat{V}(\hat{r}_{ay}) = \frac{(1 - \hat{r}_{ay}^2)^2}{2} \frac{\hat{s}(\hat{h}_{ax}^2)\hat{s}(\hat{h}_{ay}^2)}{\hat{h}_{ax}^2 \hat{h}_{ay}^2}, \text{ em que } \hat{s}(\hat{h}_{ax}^2) = [\hat{V}(\hat{h}_{ax}^2)]^{1/2}.$$

No presente caso:

$$\hat{V}(\hat{h}_{ay}^2) = \frac{32.0,3147}{4.3.62} = 0,0135 \text{ e } \hat{s}(\hat{h}_{ay}^2) = 0,1163 \text{ para o peso de frutos;}$$

$$\hat{V}(\hat{h}_a^2) = \frac{32.0,2647}{4.3.62} = 0,0114 \text{ e } \hat{s}(\hat{h}_a^2) = 0,1067 \text{ para o diâmetro da copa;}$$

$$\hat{V}(\hat{r}_{ay}) = \frac{[1 - (-0,2832)^2]^2}{2} \frac{0,1067.0,1163}{0,2647.0,3147} = 0,0630;$$

$$\hat{s}(\hat{r}_{ay}) = 0,2510.$$

Dessa forma, no presente caso, estimativas para os parâmetros genéticos são:

$$\hat{h}_{ax}^2 = 0,2674 \pm 0,1067, \text{ ou } 0,158 - 0,3714 \text{ a } 67\% \text{ de confiança (t de Student = 1);}$$

$$\hat{h}_{ay}^2 = 0,3147 \pm 0,1163, \text{ ou } 0,1984 - 0,4310 \text{ a } 67\% \text{ de confiança;}$$

$$\hat{r}_{ay} = -0,2832 \pm 0,2510, \text{ ou } -0,5342 - 0,0322 \text{ a } 67\% \text{ de confiança.}$$

Verifica-se que os parâmetros não foram estimados com muita precisão, provavelmente em função da amostragem restrita. Devido à esta baixa precisão e dos aspectos relacionados à robustez dos métodos de seleção, descarta-se a possibilidade da seleção bivariada. Por outro lado, a seleção pelo índice multiefeitos univariada pode ser aplicada, não conduzindo, provavelmente, a erros na seleção mas, implicando erros nas estimativas do progresso genético.

Resende & Rosa-Perez (1999a) relatam que desvios padrões da ordem de até 20% do verdadeiro valor da herdabilidade seriam aceitáveis para uso na predição de valores genéticos. Neste caso, seriam necessários os seguintes números totais (npb) de indivíduos nos experimentos: 8.000, 4.000 e 2.600 para herdabilidades de 0,10; 0,20 e 0,30, respectivamente. Tomando por base o limite inferior (0,1984) da estimativa da herdabilidade para peso de frutos, pode-se inferir que, na próxima geração (avaliação de 103 famílias), a utilização de 10 repetições (40 plantas/família) é recomendada para atingir uma boa precisão (cerca de 4.000 indivíduos) na estimativa da herdabilidade. Este número (nb=40) já foi também recomendado tendo em vista a maximização da acurácia pelo índice multiefeitos. Tal número de indivíduos é também adequado à estimativa da correlação genética, desde que as herdabilidades sejam aproximadamente iguais, assim como as correlações genética e ambiental. Pela expressão da variância de r_{axy} , pode-se inferir que caracteres com altas herdabilidades e baixos desvios padrões destas herdabilidades tendem a gerar estimativas mais precisas de correlação genética.

Uma formulação alternativa para o cômputo do desvio padrão da herdabilidade individual no sentido restrito, estimada a partir de testes de progênies de meios-irmãos é dada por Becker (1984):

$$\hat{s}(\hat{h}_a^2) = \left[16 * \hat{V}(\hat{\sigma}_p^2) / (\sigma_f^2)^2 \right]^{1/2}, \text{ em que:}$$

$\hat{V}(\hat{\sigma}_p^2) = [2/b^2][Q_2^2/(gl_2 + 2) - Q_3^2/(gl_3 + 2)]$: estimativa da variância do componente de variância entre progênies;

$\hat{\sigma}_f^2 = \hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_d^2$: estimativa da variância fenotípica individual;

gl_2 e gl_3 = graus de liberdade associados aos quadrados médios de progênies (Q_2) e do resíduo (Q_3).

No presente exemplo, para o caráter produção de frutos:

$$\hat{V}(\hat{\sigma}_p^2) = 13,4841;$$

$$\hat{\sigma}_f^2 = 127,0466;$$

$$\hat{s}(\hat{h}_a^2) = 0,1156.$$

Verifica-se que, neste caso, o valor obtido de $\hat{s}(\hat{h}_a^2)$ é, praticamente, equivalente àquele (0,1163) obtido empregando-se a expressão apresentada por Falconer (1989).

Existem métodos mais precisos para obtenção dos desvios padrões das estimativas da herdabilidade e do ganho genético. Estes métodos baseiam-se nos processos de reamostragem (Bootstrap e Jackknife), os quais utilizam-se dos próprios dados amostrais e não de dados simulados. Estes métodos são bem descritos por Efron (1982) e Manly (1997) e são apresentados com mais detalhe no Capítulo 21. No método Bootstrap, a autocorrelação entre subamostras é reduzida e a proporção de dados comuns às subamostras é baixa. Tal método vem sendo muito utilizado em genética de populações, cuja estruturação dos dados é mais simples.

O método Jackknife processa-se por meio de reamostragens exaustivas, gerando subamostras fortemente autocorrelacionadas, uma vez que opera sobre subamostras cujo número de membros é próximo daquele da amostra total. Assim, é indicado para situações e estruturas mais complexas e frágeis. É especialmente apropriado para a estimação dos desvios padrões da herdabilidade e do ganho genético. Entretanto, tal método deve ser adaptado para remover vícios devidos à autocorrelação ou correlação serial, adotando-se, por exemplo o método de Quenouille-Tukey para a estimação não viciada da variância do estimador. O método robusto de Jackknife é bem descrito por Lebart et al. (1979). O estimador Jackknife para o erro padrão de $\hat{\theta}$ é $\hat{\sigma}_{jk} / \sqrt{n}$ em que

$$\hat{\sigma}_{jk}^2 = (n-1) \sum_{i=1}^n (\hat{\theta}_i - \hat{\theta}_{(.,n)})^2, \quad \hat{\theta}_{(.,n)} = \sum_{i=1}^n \theta_i / n \text{ e } \hat{\theta}_i \text{ é o estimador de } \theta \text{ computado com a i-ésima observação removida (Buzas, 1997).}$$

5.2. Testes Clonais

Considere a avaliação de $c=41$ clones de café da espécie robusta (caráter produção de frutos), no delineamento em blocos ao acaso, com $b=3$ repetições e $n=5$ plantas por parcela. A análise de variância apresentou os seguintes resultados:

FV	QM	QM	E(QM)
Blocos	Q_1	0,00148	-
Clones	Q_2	0,00579	$\sigma_d^2 / n + \sigma_e^2 + b\sigma_g^2$
Resíduo	Q_3	0,00054	$\sigma_d^2 / n + \sigma_e^2$
Dentro de Parcela	Q_4	0,00177	σ_d^2
Média = 0,1321			

A partir dos resultados da análise de variância, são encontradas as seguintes estimativas:

$$\hat{\sigma}_g^2 = \frac{Q_2 - Q_3}{b} = 0,00175 = \text{variância genotípica entre clones};$$

$$\hat{\sigma}_e^2 = Q_3 - Q_4/n = 0,000186 = \text{variância entre parcelas};$$

$$\hat{\sigma}_d^2 = Q_4 = 0,00177 = \text{variância ambiental dentro de parcelas};$$

$$\hat{\sigma}_f^2 = \hat{\sigma}_g^2 + \hat{\sigma}_e^2 + \hat{\sigma}_d^2 = 0,003706 = \text{variância fenotípica};$$

$$\hat{h}_g^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_f^2} = 0,4722 = \text{herdabilidade individual no sentido amplo};$$

$$\hat{c}^2 = \hat{\sigma}_e^2 / \hat{\sigma}_f^2 = 0,0502 = \text{correlação devida ao ambiente comum de parcela.}$$

Neste caso, a única modalidade de seleção pertinente é aquela baseada no efeito de clones dado por $(\bar{Y}_i.. - \bar{Y}...)$. Assim, para a predição dos efeitos genotípicos, tem-se $I = b_{gf}(\bar{Y}_i.. - \bar{Y}...)$.

$$b_{gf} = \frac{\text{cov}[g, (\bar{Y}_i.. - \bar{Y}...)]}{\text{Var}(\bar{Y}_i.. - \bar{Y}...)} = \text{regressão do valor genotípico sobre o efeito de clone ou herdabilidade da média de clone.}$$

$$\text{cov}[g, (\bar{Y}_i.. - \bar{Y}...)] = \sigma_g^2 \frac{c-1}{c}$$

$$\text{Var}(\bar{Y}_i.. - \bar{Y}...) = (\sigma_g^2 + \sigma_e^2/b + \sigma_d^2/nb) \frac{c-1}{c}$$

$$b_{gf} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/b + \sigma_d^2/nb} = \frac{nbh_g^2}{1 + (n-1)(h_g^2 + c^2) + (b-1)nh_g^2}$$

No presente caso:

$$\hat{b}_{gf} = \frac{5.3.0,4722}{1 + (5-1)(0,4722 + 0,0502) + (3-1).5.0,4722} = \frac{7,0830}{7,8116} = 0,9067 \cong 0,91$$

Dessa forma, o melhor clone apresenta média fenotípica igual a 0,2352 e o seu valor genotípico predito equívale a:

$$\hat{g} = \hat{I} = 0,9067(0,2352 - 0,1321) + 0,1321 = 0,2256$$

Acurácia seletiva é dada por

$$\hat{r}_{lg} = [\hat{b}_{gf}]^{1/2} = 0,9522$$

O intervalo de confiança do valor genético predito é dado por

$$\hat{g} \pm t[(1 - \hat{r}_{lg}^2)\hat{\sigma}_g^2]^{1/2} = 0,2256 \pm 0,02505 = 0,2006 - 0,2506$$

Em experimentos com uma planta por parcela, a herdabilidade da média do clone é dada por:

$$b_{gf} = \frac{bh_g^2}{1 + (b-1)h_g^2} = \frac{b}{b + \frac{1-h_g^2}{h_g^2}} = \frac{b}{b + \alpha}, \text{ em que } \alpha = \frac{1-h_g^2}{h_g^2}.$$

Neste caso, a acurácia é dada por $\hat{r}_{lg} = \left[\frac{b}{b + \alpha} \right]^{1/2}$.

No presente caso, $h_g^2 = 0,4722$ e $\alpha = 1,1177$ e, portanto, a acurácia equívale a $\hat{r}_{lg} = [b/(b+1,1177)]^{1/2}$. Assim, ter-se-iam as seguintes acuráncias em função do número b de ramegos por clone: 0,9040; 0,9484; 0,9647 e 0,9732 para b=5; 10; 15 e 20, respectivamente. Assim, com uma planta por parcela, seria suficiente plantar apenas 10 ramegos por clone para se atingir uma acurácia de 0,95.

A variância aproximada da estimativa da herdabilidade no sentido amplo é dada por $Var(h_g^2) = 8h_g^2 / (nbc)$ (Resende & Rosa-Perez, 1999a). Para obter um desvio padrão de h_g^2 , da ordem de 20% de h_g^2 , tem-se, no presente caso:

$$\hat{s}(h_g^2) = [8h_g^2 / (nbc)]^{1/2}$$

$$[0,20 \cdot 0,4722]^2 = 8,04722 / (nbc)$$

$nbc = 423,55 \cong 424$ indivíduos.

Dessa forma, com nb=10 ramegos por clone, seria necessária a avaliação de 43 clones. Com c=41 clones, seria necessário utilizar nb=11 ramegos por clone.

Na situação real, com c=41 e nb=15 ramegos por clone $\hat{s}(h_g^2) = [8,04722 / 615]^{1/2} = 0,0784$ e $\hat{h}_g^2 = 0,4722 \pm 0,0784$, ou seja $\hat{s}(h_g^2) = 0,17h_g^2$. Conclui-se, então, que o número de ramegos por clone (nb) e o número total de ramegos (nbc), empregados foram adequados, tanto para a estimativa de parâmetros genéticos quanto para propiciar uma alta acurácia seletiva.

5.3. Medidas Repetidas

Nas situações em que somente a “seleção massal estratificada baseada em medidas repetidas” puder ser utilizada, uma modelagem bem mais simples pode ser adotada, conforme será detalhado a seguir, com base em Resende & Silva (1991) e Resende (2000).

A título de ilustração, considere o exemplo a seguir, referente à avaliação de três coqueiros em duas safras, para o caráter número de frutos. Os números de frutos na safra 1 foram 10, 8 e 9, para os coqueiros C₁, C₂ e C₃, respectivamente. Na safra 2, as produções foram 15, 7 e 12, para C₁, C₂ e C₃, respectivamente.

mente. Empregando-se o método da análise de variância para a estimativa da repetibilidade, tem-se:

Obtenção das Somas de Quadrados

Plantas (P)	C ₁	C ₂	C ₃	ΣM_i
Medições (M)				
M ₁	10	8	9	27
M ₂	15	7	12	34
ΣP_i	25	15	21	61

$$\text{Correção} = C = \frac{(\sum y_i)^2}{n} = \frac{61^2}{6} = 620,1667$$

SQTotal

$$= \sum y_i^2 - \frac{(\sum y_i)^2}{n} = (10^2 + 8^2 + 9^2 + 15^2 + 7^2 + 12^2) - C = 663 - C = 42,8333$$

$$\text{SQM} = \frac{27^2 + 34^2}{3} - C = \frac{1885}{3} - C = 8,1666$$

$$\text{SQP} = \frac{25^2 + 15^2 + 21^2}{2} - C = \frac{1291}{2} - C = 645,5 - 620,1667 = 25,3333$$

$$\text{SQ}(M \times P) = \text{SQTotal} - \text{SQM} - \text{SQP} = 42,8333 - 8,1666 - 25,3333 = 9,3334.$$

Análise de variância

FV	GL	SQ	QM	E(QM)
Medições (M)	1	8,1666	$Q_1 = 8,1666$	-
Plantas (P)	2	25,3333	$Q_2 = 12,6667$	$\sigma_{et}^2 + m\sigma_p^2$
M x P	2	9,3334	$Q_3 = 4,6667$	σ_{et}^2
Total	5	42,8333		

A partir desta análise, têm-se as estimativas:

$$\hat{\sigma}_p^2 = \frac{Q_2 - Q_3}{m} = \frac{12,6667 - 4,6667}{2} = 4,0000: \text{ estimativa da variância permanente entre plantas;}$$

$\hat{\sigma}_{et}^2 = Q_3 = 4,6667$: estimativa da variância ambiental temporária;

$\hat{\rho} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{et}^2} = \frac{4,0000}{4,0000 + 4,6667} = 0,4615$: estimativa do coeficiente de repetibilidade individual;

$\hat{\sigma}_f^2 = \hat{\sigma}_p^2 + \hat{\sigma}_{et}^2 = 8,6667$: estimativa da variância fenotípica individual.

O componente $\hat{\sigma}_p^2$ estima toda a variância genotípica σ_g^2 mais a variância do efeito de ambiente permanente σ_{ep}^2 , efeito este peculiar a cada planta e que surge de fatores ambientais (por exemplo, estrutura e fertilidade do solo na cova de plantio) que começam a atuar após o plantio e que permanecem durante as várias (m) safras, como um efeito permanente. Por outro lado, σ_{et}^2 representa a variação ambiental temporária associada aos efeitos ambientais efêmeros que se manifestam em cada medição, tais quais os efeitos das flutuações climáticas ano a ano e suas interações com os efeitos de plantas. Dessa forma, tem-se $\sigma_p^2 = \sigma_g^2 + \sigma_{ep}^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2$, em que σ_a^2 e σ_d^2 referem-se à variância genética aditiva e variância genética de dominância, respectivamente.

O coeficiente de repetibilidade pode então ser escrito como:

$$\rho = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{et}^2} = \frac{\sigma_g^2 + \sigma_{ep}^2}{\sigma_f^2} = \frac{\sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2}{\sigma_f^2}$$

Têm-se, ainda, as seguintes definições paramétricas:

$h_g^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_f^2} = \frac{\sigma_g^2}{\sigma_f^2}$: herdabilidade individual no sentido amplo;

$h_a^2 = \frac{\sigma_a^2}{\sigma_f^2}$: herdabilidade individual no sentido restrito.

Por estas expressões verifica-se que a repetibilidade fornece o limite superior dos coeficientes de herdabilidade. O coeficiente ρ , h_g^2 e h_a^2 são os coeficientes de determinação dos valores fenotípico permanente (p), genotípico (g) e genético aditivo (a), respectivamente, com base na seleção pelo valor fenotípico em uma só medição. Tais valores podem também ser preditos, considerando uma seleção baseada em médias de safras ou medições, pelas expressões:

- (a) Valor fenotípico permanente ou valor esperado de uma produção futura ou capacidade provável de produção (\hat{p})

$\hat{p} = MG + \beta_p (MP_i - MG)$, em que:

$$\beta_p = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{et}^2 / m} = \frac{m\hat{\rho}}{1 + (m-1)\hat{\rho}} = \frac{2 \cdot 0,4615}{1 + (2-1) \cdot 0,4615} = 0,6315: \text{ coeficiente de determinação do valor fenotípico permanente ou repetibilidade da média de } m = 2 \text{ medições no indivíduo;}$$

MG : média geral das plantas nas várias medições = 10,17;
 MP_i : média geral da planta i nas várias medições, equivalendo a 12,5; 7,5 e 10,5, para as plantas 1, 2 e 3, respectivamente.

No presente exemplo, os valores fenotípicos permanentes preditos são $\hat{p}_1 = 11,64$; $\hat{p}_2 = 8,48$ e $\hat{p}_3 = 10,38$ para as plantas 1, 2 e 3, respectivamente.

(b) Valor genotípico (\hat{g})

$\hat{g} = MG + \beta_g (MP_i - MG)$, em que:

$$\beta_g = \frac{m\hat{h}_g^2}{1 + (m-1)\hat{\rho}}: \text{ coeficiente de determinação do valor genotípico real do indivíduo ou herdabilidade no sentido amplo da média de } m \text{ medições no indivíduo;}$$

Assumindo $\hat{h}_g^2 = 0,30$, tem-se $\beta_g = 0,4105$ para $m = 2$ medições. Neste caso, os valores genotípicos preditos das 3 plantas são $\hat{g}_1 = 11,13$; $\hat{g}_2 = 9,07$ e $\hat{g}_3 = 10,31$.

(c) Valor genético aditivo (\hat{a})

$\hat{a} = MG + \beta_a (MP_i - MG)$, em que:

$$\beta_a = \frac{m\hat{h}_a^2}{1 + (m-1)\hat{\rho}}: \text{ coeficiente de determinação do valor genético aditivo ou herdabilidade no sentido restrito da média de } m \text{ medições no indivíduo.}$$

Considerando $\hat{h}_a^2 = 0,20$, tem-se $\beta_a = 0,2736$ para $m = 2$ medições. Neste caso, os valores genéticos aditivos preditos das 3 plantas são $\hat{a}_1 = 10,81$; $\hat{a}_2 = 9,43$ e $\hat{a}_3 = 10,26$.

Das três quantidades apresentadas, em geral, \hat{p} não apresenta utilidade prática pois representa o valor da planta naquele microambiente da cova onde está plantada. Por outro lado, \hat{g} é o valor esperado da planta se a mesma for utilizada para plantios por via vegetativa ou assexuada e $(1/2)\hat{a}$ é a contribuição da planta para seu descendente por via sexuada. Tomando como exemplo a planta 1 têm-se as seguintes previsões: valor fenotípico médio: 12,5 frutos; valor fenotípico permanente: 11,64 frutos; valor genotípico: 11,13 frutos; valor genético aditivo: 10,81 frutos. Com base nesses valores e nos estimadores de

ρ , h_g^2 e h_a^2 , verifica-se que o valor fenotípico permanente é um preditor mais adequado para o valor genotípico do que para o valor genético aditivo. Assim, na caracterização e avaliação de germoplasmas recomenda-se a apresentação de Tabelas com os valores fenotípicos permanentes ou os valores genotípicos preditos e não os valores fenotípicos médios. Dessa forma, ter-se-ia uma boa inferência sobre a produtividade do germoplasma, quando utilizado via clonagem.

As acurárias nas predições de p , g e a são dadas por $[\beta_p]^{1/2}$; $[\beta_g]^{1/2}$ e $[\beta_a]^{1/2}$. Assim, constata-se pelas expressões de β_p , β_g e β_a que quanto maior o número m de medições em cada planta, maiores são as acurárias das predições e, portanto, mais precisas são as inferências. É importante relatar que β_p refere-se ao coeficiente de determinação do valor fenotípico permanente (o qual não tem valor prático) e não do valor genotípico real. Dessa forma, considerando um caráter com $\rho = 0,65$; $h_g^2 = 0,30$ e $h_a^2 = 0,20$ e realizando-se $m = 3$ medições em cada planta tem-se que a determinação do valor fenotípico permanente (β_p) equivale a 85%, mas o coeficiente de determinação do valor genotípico (β_g) equivale apenas a 39% e o coeficiente de determinação do valor genético aditivo (β_a) equivale apenas a 26%. Os valores máximos a serem considerados (com o aumento de m) pelos coeficientes de determinação do valor genotípico real e do valor genético aditivo são dados por h_g^2 / ρ e h_a^2 / ρ , respectivamente, expressões estas obtidas por derivação parcial de β_g e β_a em relação a m . Com $\rho = 0,65$, $h_g^2 = 0,30$ e $h_a^2 = 0,20$, as determinações máximas possíveis são 46,15% e 30,77% para g e a , respectivamente.

A eficiência, em termos de ganho genético com seleção, do uso de m medições em cada planta em relação ao uso de apenas uma é dada por $[m/[1+(m-1)\rho]]^{1/2}$, expressão esta válida tanto para g (estratégia assexuada de propagação do material selecionado) quanto para a (estratégia sexuada de propagação do material selecionado). Para um caráter com $\rho = 0,65$ as eficiências são: 1,10; 1,14; 1,16; 1,18 e 1,19 para $m = 2$; 3; 4; 5 e 6 medições, respectivamente. Neste caso, praticamente não compensa avaliar mais do que três safras em cada planta. Com $m = 3$ atinge-se 85% da determinação máxima possível ($m = \infty$) para os valores genotípicos e também para os valores aditivos. O valor de m para se atingir uma fração f da determinação máxima possível é dado por

$$m = \frac{f}{(1-f)} \frac{(1-\rho)}{\rho}.$$

Para o caráter em consideração, para atingir $f = 0,90$ ou 90% da

determinação máxima possível seriam necessárias $m = 4,85 \approx 5$ medições. É importante relatar que as eficiências apresentadas são as eficiências máximas

possíveis e são válidas para a seleção massal baseada em médias de safras. Quando informações de famílias, progêneres ou de testes clonais são também utilizadas na seleção, essas eficiências pelo uso de medidas repetidas são reduzidas.

De maneira genérica, recomenda-se a estratificação do plantio em blocos e a realização da “seleção massal estratificada baseada em médias de safras”. Considere por exemplo $p = 20$ plantas distribuídas em $b = 4$ estratos ou blocos com $p/b = 5$ plantas por estrato e a realização de $m = 3$ medições por planta. Neste caso, a estimativa da repetibilidade pelo método da análise de variância segue o esquema:

FV	GL	E(QM)
Bloco-Medição	$(mb-1) = 11$	-
Planta	$(p-1) = 19$	$\sigma_{et^*}^2 + m\sigma_{p^*}^2$
Resíduo	$(p/b-1)(m-1)*b-(b-1)=29$	$\sigma_{et^*}^2$
Total	$pm - 1 = 59$	

Nesta situação, os efeitos de blocos, medições e interação bloco x medições são tomados como fixos e ajustados em uma única fonte de variação denominada combinação bloco-medição, a qual contempla o somatório dos graus de liberdade dos referidos efeitos. A fonte de variação denominada “resíduo” pode ser obtida facilmente por diferença em relação à soma de quadrados referentes a “total”. O componente de variância $\sigma_{p^*}^2$ deve apresentar magnitude inferior a σ_p^2 obtido sem estratificação, visto que a variância entre blocos, a qual superestima a variância de ambiente permanente (σ_{ep}^2) em σ_p^2 , é removida. Com a redução de σ_{ep}^2 , $\hat{\rho}$ se aproxima mais de h_g^2 , de forma que o valor fenotípico permanente torna-se um melhor preditor do valor genotípico.

No caso da “seleção massal estratificada baseada em médias de safras”, o ordenamento dos indivíduos para fins de seleção deve basear-se no seguinte preditor dos valores genéticos: $\hat{g} = MG + \beta_g (MP_i - MB_i)$, em que MB_i refere-se à média do bloco em que encontra-se a planta i . Esta expressão é válida quando o número m de medições é o mesmo para todas as plantas.

Não se conhecendo β_g , pode-se usar β_p em seu lugar sem afetar o ordenamento para fins de seleção.

Em resumo, nos casos em que a seleção massal for necessária, como, por exemplo, sobre materiais estabelecidos em bancos de germoplasma sem a estrutura de família ou em campos de produtores, recomenda-se praticar a

“seleção massal estratificada baseada em medidas repetidas”, utilizando estimativas de repetibilidade e herdabilidades (estimativas prévias) nos sentidos amplo e restrito para inferências sobre os valores genotípicos e genéticos aditivos.

Capítulo 6

**Repetibilidade e Avaliação
Genética com Medidas Repetidas**



6.1. Conceitos

A realização de medidas repetidas (no tempo ou no espaço) em um caráter no indivíduo é prática comum no melhoramento de espécies perenes como a erva-mate (Resende & Silva, 1991; Resende et al., 1995a), a seringueira (Gonçalves et al., 1990), o coqueiro (Siqueira, 1982), o cacaueiro (Dias & Souza, 1993; Resende & Dias, 2000), o cupuaçuzeiro (Fonseca et al., 1990; Costa et al., 1997), o guaranazeiro (Valois et al., 1979) e, também, em espécies florestais, em associação com medidas incrementais de crescimento (Resende, 1999b).

A avaliação genética de indivíduos destas espécies deve, então, basear-se em modelos que consideram o efeito adicional denominado efeito de ambiente permanente, bem como o parâmetro associado à correlação fenotípica entre medidas repetidas em um mesmo indivíduo, o qual é chamado repetibilidade.

O coeficiente de repetibilidade mede a capacidade dos organismos em repetir a expressão do caráter, ao longo de vários períodos de tempo (ou no espaço), no decorrer de suas vidas. Do ponto de vista prático, este parâmetro apresenta importância fundamental na predição de valores genéticos e genotípicos e na inferência sobre o aumento da eficiência seletiva pelo uso de um determinado número ($>$ que 1) de medições por indivíduo, fator que permite determinar o número de safras a ser adotado em um programa prático de melhoramento genético.

A repetibilidade foi conceituada por Lush (1937) como correlação fenotípica intraclasse e, portanto, no caso de indivíduos, refere-se às correlações fenotípicas entre medições repetidas no mesmo indivíduo. Esta é a repetibilidade individual. Outro conceito é da repetibilidade maternal o qual apresenta maior relevância para o melhoramento animal e não será abordado neste documento. Detalhes relativos à repetibilidade maternal são apresentados em Resende & Rosa-Perez (1999a) e Resende & Freitas (1998).

Uma condição essencial para a adoção do modelo de repetibilidade é que as medidas repetidas apresentem correlação genética igual ou próxima a 1, fato que confirma ser o mesmo caráter de uma medição para outra. Caso contrário, deve-se adotar o modelo multivariado, o qual trata cada medição como um caráter distinto. Neste documento, as abordagens multivariadas para a análise de medidas repetidas são apresentadas nos Capítulos 7, 9 e 11, nos quais os efeitos genéticos são tratados como aleatórios e, no Capítulo 13, no qual os efeitos genéticos são tratados como fixos.

6.2. Repetibilidade e Predição de Valores Genéticos, Genotípicos e Fenotípicos Permanentes

Considerando um caráter, como o peso de frutos de dendê, tem-se o seguinte modelo (ignorando a epistasia) para as observações fenotípicas em um dado estágio:

$y = \mu + a + d + e_p + e_t$, em que:

y : valor fenotípico temporário.

μ : média genotípica ou média geral.

a : efeito aditivo.

d : efeito de dominância.

e_p : efeito de ambiente permanente.

e_t : efeito de ambiente temporário.

$\mu + a$: valor genético aditivo.

$\mu + a + d = \mu + g$: valor genotípico.

$\mu + a + d + e_p$: valor fenotípico permanente.

Ao praticar a seleção, o melhorista de plantas perenes pode ter como alvo três tipos de população melhorada de referência formada pelos: (i) descendentes dos indivíduos selecionados; (ii) próprios indivíduos selecionados, propagados vegetativamente e (iii) próprios indivíduos selecionados, mantidos no mesmo microambiente em que estão plantados (desbaste das piores plantas). À população melhorada referenciada em (i) corresponde a população da geração seguinte à seleção e as populações melhoradas em (ii) e (iii) correspondem à população na mesma geração plantada em outro ambiente e mantida no mesmo ambiente de seleção, respectivamente.

Assim, é relevante a predição dos valores genéticos aditivos, genotípicos e fenotípicos permanentes, para atender às populações-alvo apresentadas em (i), (ii) e (iii), respectivamente. Isto porque somente os efeitos aditivos dos alelos são transmitidos à descendência por reprodução sexuada (população de referência (i)) e porque na manutenção dos indivíduos selecionados no próprio ambiente de seleção, visando quantificar as produções futuras destes mesmos indivíduos (população de referência (iii)), capitalizam-se os efeitos aditivos, de dominância e de ambiente permanente. Por outro lado, na população referenciada em (ii), capitalizam-se os efeitos aditivo e de dominância.

Os valores genéticos (V_a), valores genotípicos (V_g) e valores fenotípicos permanentes (V_{fp}) são variáveis aleatórias desconhecidas e devem ser preditas a partir do conhecimento de y . De maneira genérica, o princípio da regressão linear deve ser empregado para estimação de V_a , V_g e V_{fp} , ou seja, a técnica de regressão permite inferir em que proporção, y explica ou estima V_a , V_g e V_{fp} . Assim:

$$V_a = \mu + \beta_{ay} (y - \mu)$$

$$V_g = \mu + \beta_{gy} (y - \mu)$$

$V_{fp} = \mu + \beta_{fpy} (y - \mu)$, onde β_{ay} é o coeficiente de regressão do valor genético aditivo sobre o valor fenotípico y , β_{gy} é o coeficiente de regressão do valor geno-

típico sobre o valor fenotípico e β_{fpy} é o coeficiente de regressão do valor fenotípico permanente sobre o valor fenotípico temporário.

Pelo princípio da regressão os estimadores são:

$$\hat{\beta}_{ay} = \frac{COV[(\mu+a), y]}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_y^2} = h_a^2$$

$$\hat{\beta}_{gy} = \frac{COV[(\mu+g), y]}{\sigma_y^2} = \frac{\sigma_g^2}{\sigma_y^2} = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_y^2} = h_g^2$$

$$\hat{\beta}_{fpy} = \frac{COV[(\mu+g), y]}{\sigma_y^2} = \frac{\sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2}{\sigma_y^2} = \rho, \text{ em que:}$$

h_a^2 , h_g^2 e ρ : herdabilidades individuais no sentido restrito e amplo e repetibilidade, respectivamente.

$$\sigma_y^2 = Var(\mu+a+d+ep+et) = \sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2 + \sigma_{et}^2$$

σ_a^2 , σ_d^2 , σ_{ep}^2 e σ_{et}^2 : variâncias genética aditiva, de dominância, ambiental permanente e ambiental temporária, respectivamente.

Os estimadores de β apresentados são adequados apenas à seleção individual com uma medição por indivíduo.

Verifica-se que o coeficiente de repetibilidade representa o limite superior dos coeficientes de herdabilidade no sentido amplo (associado à propagação clonal) e restrito (associado à propagação por sementes). Assim, pouco pode ser dito sobre os coeficientes de herdabilidade quando o coeficiente de repetibilidade é alto. De maneira geral, o coeficiente de repetibilidade é mais próximo do coeficiente de herdabilidade no sentido amplo do que do coeficiente de herdabilidade no sentido restrito. Assim, o uso da repetibilidade como um parâmetro isolado na predição é mais útil para atender as populações-alvo melhoradas em (iii) (apenas eliminação das piores plantas de um plantio) ou (ii) (plantio clonal de indivíduos selecionados) (Resende & Silva, 1991). O uso apenas da repetibilidade para inferências (predições) sobre o valor genético aditivo (plantio por sementes) é limitado.

Se medições repetidas são realizadas sobre cada indivíduo, os preditores de a , g e $(g+e_p)$, a partir do valor fenotípico médio de várias medições (\bar{y}) são dados por:

$$\hat{a} = \hat{\beta}_{a\bar{y}} (\bar{y} - \mu) = \frac{COV(a, \bar{y})}{Var(\bar{y})} (\bar{y} - \mu) = \frac{mh_a^2}{1+(m-1)\rho} (\bar{y} - \mu)$$

$$\hat{g} = \hat{\beta}_{g\bar{y}} (\bar{y} - \mu) = \frac{COV(g, \bar{y})}{Var(\bar{y})} (\bar{y} - \mu) = \frac{mh_g^2}{1+(m-1)\rho} (\bar{y} - \mu)$$

$$(\hat{g} + \hat{e}_p) = \hat{\beta}_{(g+ep), \bar{y}} (\bar{y} - \mu) = \frac{COV[(g+e_p), \bar{y}]}{Var(\bar{y})} (\bar{y} - \mu) = \frac{m\rho}{1+(m-1)\rho} (\bar{y} - \mu)$$

Verifica-se por estas expressões que o ganho em eficiência pelo uso de m medições por indivíduo, em vez de uma, é dado por $G_{s_m}/G_{s_1} = r_{\hat{a}am}/r_{\hat{a}a1} = \{m/[1+(m-1)\rho]\}^{1/2}$, em que G_{s_m} é o ganho genético com m medições por indivíduo, G_{s_1} é o ganho genético com uma medição por indivíduo, $r_{\hat{a}am}$ é a acurácia seletiva com m medições e $r_{\hat{a}a1}$ é a acurácia seletiva com uma medição por indivíduo. Embora o desenvolvimento desta expressão tenha se baseado nos efeitos aditivos, tal expressão se presta, também, aos efeitos genotípicos e fenotípicos permanentes.

6.3. Repetibilidade e Predição de Valores Genéticos e Genotípicos em Testes de Progêneres e Clonais

No caso de testes de progêneres, o índice multiefeitos para a predição dos valores genéticos aditivos depende dos parâmetros h_a^2 e ρ e é dado por (Resende et al., 2000a):

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

$$b_1 = \frac{(1-\rho_a) m h_a^2}{1+(m-1)\rho - m [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2) + c^2]} = \frac{0,75 m h_a^2}{1+(m-1)\rho - m (0,25 h_a^2 + c^2)}, \text{ para famílias de meios-irmãos.}$$

$$\begin{aligned} b_2 &= \frac{[1+(nb-1)\rho_a] m h_a^2}{1+(m-1)\rho + (n-1)m[\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2) + c^2] + (b-1)nm[\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]} \\ &= \frac{[1+(nb-1)0,25] m h_a^2}{1+(m-1)\rho + (n-1)m(0,25 h_a^2 + c^2) + (b-1)nm0,25 h_a^2}, \text{ para famílias de meios-irmãos.} \end{aligned}$$

$$\begin{aligned} b_3 &= \frac{[1-\rho_a] m h_a^2}{1+(m-1)\rho + (n-1)m[\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2) + c^2] - nm[\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]} \\ &= \frac{0,75 m h_a^2}{1+(m-1)\rho + (n-1)m(0,25 h_a^2 + c^2) - nm0,25 h_a^2}, \text{ para famílias de meios-irmãos.} \end{aligned}$$

m = número de medições (avaliações) por indivíduo.

$\bar{y}_{ijk..}$, $\bar{y}_{ij..}$, $\bar{y}_{i..}$ e $\bar{y}_{...}$ = médias do indivíduo, da parcela, da família, do bloco e geral, respectivamente.

Verifica-se pelas fórmulas de b_1 , b_2 e b_3 , que a realização de várias medições (m) por indivíduo contribui para minimizar (desde que a repetibilidade não seja 1) as variâncias fenotípicas ao nível de médias e, consequentemente, para maximizar a eficiência seletiva.

O índice multiefeitos para a predição dos efeitos genotípicos (visando à propagação assexuada) depende de h_a^2 , h_g^2 e ρ e é dado por:

$$I = \hat{g} = (b_1 + b_4) (\bar{y}_{ijk} - \bar{y}_{ij.}) + (b_2 + b_5) (\bar{y}_{i...} - \bar{y}_{....}) + (b_3 + b_6) (\bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{j..} + \bar{y}_{....}), \text{ em que:}$$

$$b_4 = \frac{(1-\rho_d) m (h_g^2 - h_a^2)}{1+(m-1) \rho - m [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2) + c^2]} = \frac{m (h_g^2 - h_a^2)}{1+(m-1) \rho - m (0,25 h_a^2 + c^2)}, \text{ para famílias de meios-irmãos;}$$

$$\begin{aligned} b_5 &= \frac{[1+(nb-1)\rho_d]m (h_g^2 - h_a^2)}{1+(m-1) \rho + (n-1) m [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2) + c^2] + (b-1) nm [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]} \\ &= \frac{m (h_g^2 - h_a^2)}{1+(m-1) \rho + (n-1) m (0,25 h_a^2 + c^2) + (b-1) nm 0,25 h_a^2}, \text{ para famílias de meios-irmãos;} \end{aligned}$$

$$\begin{aligned} b_6 &= \frac{(1-\rho_d) m (h_g^2 - h_a^2)}{1+(m-1) \rho + (n-1) m [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2) + c^2] - nm [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)]} = \\ &= \frac{m (h_g^2 - h_a^2)}{1+(m-1) \rho + (n-1) m (0,25 h_a^2 + c^2) - nm 0,25 h_a^2}, \text{ para famílias de meios-irmãos.} \end{aligned}$$

A seleção de matrizes com base no comportamento de suas progêniess de meios-irmãos deve ser baseada nos valores genéticos preditos por:

$$I = \hat{a} = b_7 (\bar{y}_{i..} - \bar{y}_{....})$$

$$b_7 = \frac{(1/2) nb m h_a^2}{1+(m-1) \rho + (n-1) m (0,25 h_a^2 + c^2) + (b-1) nm 0,25 h_a^2}$$

No caso de testes clonais, a predição dos efeitos genotípicos dos clones deve ser realizada pelo índice:

$$I = \hat{g} = b_8 (\bar{y}_{i...} - \bar{y}_{....}), \text{ em que:}$$

$$b_8 = \frac{n b m h_g^2}{1+(m-1) \rho + (n-1) m (h_g^2 + c^2) + (b-1) nm h_g^2} = \text{herdabilidade no sentido amplo ao nível de médias de clones.}$$

Preditores (univariados e multivariados) de valores fenotípicos permanentes a partir de testes de progêñies (índice multiefeitos para valores fenotípicos permanentes) foram apresentados por Resende & Rosa-Perez (1999a) e não serão apresentados aqui devido à menor importância para plantas do que para animais.

6.4. Eficiência Seletiva e Número de Medições por Indivíduo

6.4.1. Seleção individual

Para a seleção individual, tanto para os valores genéticos aditivos, genotípicos ou fenotípicos permanentes, duas abordagens podem ser empregadas visando avaliar a eficiência do uso de m medições por indivíduo, em vez de uma.

A primeira é dada pela expressão $\{m/[1+(m-1)\rho]\}^{1/2}$, apresentada anteriormente, a qual fornece a eficiência por ciclo seletivo. A eficiência em termos de ganho genético por unidade de tempo (por ano) é dada pela razão entre as acuráncias seletivas com m e uma medição multiplicada pela razão entre o intervalo de gerações (L) com uma medição e com m medições, ou seja, $(r_{aam}/r_{aa1})(L_1/L_m)$.

Para a erva-mate, por exemplo, o intervalo entre gerações é dado por $L = I + S + M + (m-1)$, em que I é a idade à primeira poda de produção ou colheita, S é o período (em anos) após a poda, necessário para produção de sementes, M é o período após a produção de sementes, necessário para a produção de mudas e m é o número de safras praticado. De maneira geral, $I = 3$ (Belingheri & Kricum, 1997), $S = 2$ e $M = 2$. Assim, atualmente, $L = 7$ para uma medição, sendo denotado por L_1 . Para m maior ou igual a 2, $L = L_1 + m - 1$ e o ganho anual é dado por $(r_{aam}/(L_1 + m - 1))$. Verifica-se, assim, que esforços para minimizar I , S e M contribuirão muito para aumento na taxa de melhoramento genético das populações de erva-mate, talvez até mais do que a própria maximização da acurácia seletiva.

Na Tabela 6.1 são apresentados os valores de eficiência seletiva por ciclo de seleção (EC) e anual (EA) para a seleção massal, considerando vários valores de repetibilidade e número de medições. A eficiência anual é dada por $[m/[1+(m-1)\rho]]^{1/2} [L_1/(L_1 + m - 1)]$, $m \geq 2$. Alternativamente, $EA = EC [L_1/(L_1 + m - 1)]$, $m \geq 2$.

Tabela 6.1. Valores de eficiência seletiva por ciclo de seleção (EC) e eficiência anual (EA) para a seleção massal baseada em médias de diferentes números (m) de safras, considerando diferentes valores de repetibilidade (ρ).

ρ	m	EC	EA*
0,90	1	1,00	1,00
	2	1,03	0,90
	3	1,04	0,80
0,85	2	1,04	0,91
	3	1,05	0,82
	2	1,05	0,92
0,80	3	1,07	0,83
	2	1,07	0,93
	3	1,10	0,85
0,75	4	1,11	0,77
	2	1,08	0,95
	3	1,12	0,87
0,70	4	1,14	0,79
	2	1,10	0,96
	3	1,14	0,89
0,65	4	1,16	0,81
	2	1,12	0,98
	3	1,17	0,91
0,60	4	1,20	0,84
	2	1,14	0,99
	3	1,20	0,93
0,55	4	1,23	0,86
	2	1,15	1,00
	3	1,22	0,95
0,50	4	1,26	0,89
	2	1,17	1,03
	3	1,26	0,98
0,45	4	1,30	0,91
	2	1,20	1,05
	3	1,29	1,00
0,40	4	1,35	0,94
	2	1,22	1,06
	3	1,33	1,03
0,35	4	1,40	0,98
	2	1,24	1,08
	3	1,37	1,06
0,30	4	1,45	1,02

* Para o cômputo de EA, considerou-se $L_1 = 7$.

Na interpretação da Tabela 6.1, devem ser considerados os diferentes objetivos da seleção: (i) para o melhoramento no curto prazo (maximização do ganho na geração atual), analisar a EC, (ii) para o melhoramento no longo prazo por seleção recorrente, verificar a EA. Considerando uma repetibilidade estimada de 0,75, verifica-se que para o melhoramento no curto prazo não compensa avaliar mais do que três safras e para o melhoramento no longo prazo por seleção massal o ideal é selecionar com base em apenas 1 safra. Devido a esta incompatibilidade, para o melhoramento da erva-mate, sugere-se utilizar o esquema proposto por Resende et al. (1997a, p.255) que permite conciliar os objetivos de melhoramento no curto e longo prazos.

Verifica-se, também, pela Tabela 6.1, que, com repetibilidade igual a 0,50, a utilização de uma ou duas safras contribui igualmente para o melhoramento no longo prazo. A utilização de duas e três safras só passa a ser vantajosa para o melhoramento no longo prazo quando a repetibilidade for igual ou inferior a 0,45 e 0,35, respectivamente. Pode-se também relatar que quanto menor for $I + S + M$ (ou menor L), menos vantajoso torna-se o uso de medidas repetidas.

A segunda abordagem é dada em função da acurácia (r_{aa}) seletiva ou determinação (confiabilidade) (r_{aa}^2) escolhida a priori. As determinações associadas às três populações-alvo de referência são dadas por:

$$(i) \quad r_{aa}^2 = \frac{m h_a^2}{1 + (m-1) \rho} : \text{para os efeitos aditivos.}$$

$$(ii) \quad r_{gg}^2 = \frac{m h_g^2}{1 + (m-1) \rho} : \text{para os efeitos genotípicos.}$$

$$(iii) \quad r_{fpfp}^2 = \frac{m \rho}{1 + (m-1) \rho} : \text{para os efeitos fenotípicos permanentes.}$$

Verifica-se que, por esta abordagem, as inferências sobre o valor de m adequado para atingir determinada confiabilidade na seleção para os efeitos aditivos e genotípicos dependem também das herdabilidades das características e não apenas da repetibilidade. As inferências sobre a confiabilidade na seleção para os valores fenotípicos permanentes (que são de menor importância em plantas) dependem apenas do parâmetro repetibilidade. Assim, tal abordagem baseada unicamente na repetibilidade é deficiente, quando os indivíduos selecionados forem propagados por sementes ou por clonagem, sendo exata apenas quando os indivíduos selecionados forem mantidos no mesmo microambiente em que estão plantados.

A expressão r_{fpfp}^2 rearranjada fornece $m = \frac{r_{fpfp}^2 (1 - \rho)}{\rho (1 - r_{fpfp}^2)}$ (Lush, 1937), da qual

obtém-se o número de medições necessário para atingir certa determinação

r_{fpfp}^2 para os valores fenotípicos permanentes reais dos indivíduos. O parâmetro $r_{\hat{f}pfp}^2$ equivale ao quadrado da correlação entre o valor fenotípico permanente real do indivíduo e o valor fenotípico médio de m medições.

O nível adequado de confiabilidade ou determinação a ser adotado depende da finalidade da seleção. Seleção de clones objetivando plantios clonais e seleção de indivíduos para estabelecimento de pomares de sementes ou pomares biparentais demandam precisão próxima de 100%. Isto decorre de que, nesses casos, um número muito pequeno de indivíduos (ou apenas o melhor) é selecionado, de forma que a seleção com determinação diferente do máximo pode conduzir à seleção do material genético errado.

Quando um grupo maior de indivíduos é selecionado (por exemplo, para compor uma população de melhoramento), uma determinação acima de 80% (acurácia acima de 89%) já é adequada. A determinação com essa magnitude implica troca na ordem dos melhores de uma safra para outra, mas o grupo dos melhores não deverá ser alterado significativamente.

Na Tabela 6.2, são apresentados os números de medições necessárias para selecionar, com segurança, indivíduos para populações de melhoramento (precisão de 80%) e para população de produção de propágulos melhorados (precisão de 95%).

Tabela 6.2. Números de medições (safras) necessárias para a seleção de indivíduos para população de melhoramento (determinação de 80%) e para populações de produção de propágulos melhorados (determinação de 95%), considerando diferentes valores de repetibilidade.

Repetibilidade (%)	Número de safras (determinação de 80%)	Número de safras (determinação de 95%)
10	36	171
20	16	76
30	9	44
40	6	29
50	4	19
60	3	13
70	2	8
80	1	5
90	1	2

Os valores apresentados na Tabela 6.2 referem-se apenas ao melhoramento no curto prazo (acurácia seletiva no ciclo atual de seleção). Para o melhoramento no longo prazo, veja-se a Tabela 6.1.

Com vistas à predição de valores genéticos aditivos, o m adequado é dado por $m = [r_{aa}^2(1-\rho)]/[h_a^2 - \rho r_{aa}^2]$ (Resende et al., 1995a). Neste caso, os valores máximos (em função de m alto) de acurácia são apresentados na Tabela 6.3, para vários valores de acurácia e herdabilidade.

Tabela 6.3. Acurácia máxima possível para várias combinações entre valores de herdabilidade e repetibilidade (fornecida pela fração $\sqrt{h_a^2/\rho}$, obtida por derivação parcial da expressão da acurácia).

Herdabilidade	Repetibilidade										Acurácia para uma medição
	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90		
0,10	1,00	0,71	0,58	0,50	0,45	0,41	0,38	0,35	0,33		0,32
0,20	-	1,00	0,82	0,71	0,63	0,58	0,53	0,50	0,47		0,45
0,30	-	-	1,00	0,87	0,77	0,71	0,65	0,61	0,58		0,55
0,40	-	-	-	1,00	0,89	0,82	0,76	0,71	0,67		0,63
0,50	-	-	-	-	1,00	0,91	0,85	0,79	0,75		0,71
0,60	-	-	-	-	-	1,00	0,93	0,87	0,82		0,77

Observa-se que, para um valor fixo de repetibilidade, a acurácia máxima aumenta com o aumento da herdabilidade (Tabela 6.3). Verifica-se, também, que valores mais baixos de repetibilidade respondem mais ao aumento do número de medições ou safras. Em outras palavras, quando a repetibilidade tende a 1, o aumento do número de medições não conduz ao aumento da acurácia. De maneira geral, o uso de medidas repetidas conduz a uma maior acurácia principalmente em situações de baixas herdabilidade e repetibilidade, simultaneamente (Tabela 6.3).

É interessante verificar que o número de medições necessárias para atingir certa fração da determinação máxima ($r_{\text{aamáx}}^2$), independe da herdabilidade do caráter. Neste caso, $m = \frac{f}{h_a^2 - \rho f} \frac{r_{\text{aamáx}}^2(1-\rho)}{r_{\text{aamáx}}^2}$, em que f é a fração desejada. Uma

vez que $r_{\text{aamáx}}^2 = \frac{h_a^2}{\rho}$, obtém-se $m = \frac{f(1-\rho)}{(1-f)\rho}$, ou seja, m independe de h^2 . É impor-

tante ressaltar que esta expressão é similar, mas não idêntica à de Lush, pois f é uma fração de $r_{\text{aamáx}}^2$ e r_{fffp}^2 é uma fração de 1. Logicamente, ambas conduzem às mesmas inferências para m. Os valores de m necessários para obtenção de 90% da acurácia máxima possível são apresentados na Tabela 6.4.

Tabela 6.4. Número de medições necessárias (m) para obter 90% da acurácia máxima (ou 81% da determinação máxima) possível, na seleção de indivíduos com base em médias de safras.

Repetibilidade	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
M	38	17	10	6	4	3	2	1	1

Na seleção por valores genotípicos (propagação clonal), o m adequado para a seleção com confiabilidade r_{gg}^2 é dado por $m = [r_{gg}^2(1-\rho)]/[h_g^2 - \rho r_{gg}^2]$. Os resultados apresentados nas Tabelas 6.3 e 6.4 podem ser extrapolados para a seleção clonal. Os clones podem ser selecionados, também, a partir de testes clonais com 1 ou m medições. Na Tabela 6.5 são apresentados resultados comparativos destes três procedimentos de seleção clonal.

Tabela 6.5. Comparação entre acurárias pelos diferentes processos de seleção clonal, para diferentes combinações de valores para repetibilidade e herdabilidade no sentido amplo (h_g^2).

Método	Acurácia	Repetibilidade	Herdabilidade*		
			0,2	0,4	0,6
1. individual baseado em várias medi- ções	$\sqrt{\frac{mh_g^2}{1+(m-1)\rho}}$	0,2	0,85	-	-
		0,4	0,66	0,93	-
		0,6	0,60	0,79	0,97
2. baseado em médias de clones, uma avaliação por ramete	$\sqrt{\frac{nh_g^2}{1+(n-1)h_g^2}}$	0,2	0,85	-	-
		0,4	0,85	0,93	-
		0,6	0,85	0,93	0,97
3. baseado em médias de clones, várias avaliações por ramete	$\sqrt{\frac{nmh_g^2}{1+(m-1)\rho+(n-1)mh_g^2}}$	0,2	0,85	-	-
		0,4	0,82	0,93	-
		0,6	0,79	0,91	0,97

* Acurácia para $n=m=10$ (método 1 e 2) e $nm = 10$ (método 3).

Observa-se que, para valores iguais de repetibilidade e herdabilidade no sentido amplo, os métodos se equivalem. Entretanto, para a mesma quantidade de esforços (mesmo número total nm de medições), o método 2 apresenta maiores acurárias. Portanto, é vantajoso enfatizar o aumento do número de repetições em detrimento do número de medições. E essa é uma regra geral, pois h_a^2 é sempre menor que ρ e, portanto, o aumento de n contribui mais para aumen-

tar a herdabilidade no sentido amplo ao nível de médias do que o aumento de m contribui para aumentar a repetibilidade ao nível de médias. Este procedimento concorre, também, para minimizar o tempo do ciclo seletivo.

6.4.2. Vários métodos de seleção e sistemas de propagação

As eficiências do uso de m medições para outros métodos de seleção, além da individual, dependem das expressões de acurácia e ganho genético, apresentadas a seguir:

(1) Propagação sexuada dos indivíduos selecionados

- (1.1) Seleção baseada em valores individuais, sem o uso da informação da família

Acurácia	Ganho genético
$r_{\hat{aa}_1} = \left[\frac{m h_a^2}{1 + (m-1) \rho} \right]^{1/2}$	$G_{s_1} = k r_{\hat{aa}_1} \sigma_a$

- (1.2) Seleção de genitores baseada no comportamento da progênie de meios-irmãos

Acurácia	Ganho genético
$r_{\hat{aa}_2} = 0,5 \left[\frac{m N h_a^2}{1 + (m-1) \rho + (N-1) m 0,25 h_a^2} \right]^{1/2}$	$G_{s_2} = k r_{\hat{aa}_2} \sigma_a$

- (1.3) Seleção baseada em valores individuais e na informação da família de meios-irmãos

Acurácia

$$r_{\hat{aa}_3} = \left[\frac{0,5625 m \frac{N-1}{N} h_a^2}{1 + (m-1) \rho - m 0,25 h_a^2} + \frac{m [1 + (N-1) 0,25]^2 h_a^2}{N [1 + (m-1) \rho + (N-1) m 0,25 h_a^2]} \right]^{1/2}$$

Ganho genético

$$G_{s_3} = k r_{\hat{aa}_3} \sigma_a$$

(1.4) Seleção baseada em valores individuais e na informação da família de irmãos germanos

Acurácia

$$r_{aa} = \left[\frac{0,25 m \frac{N-1}{N} h_a^2}{1+(m-1) \rho - m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]} + \frac{m [1+(N-1) 0,5]^2 h_a^2}{N [1+(m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]]} \right]^{1/2}$$

Ganho genético

$$G_{s_4} = k r_{aa_4} \sigma_a$$

(1.5) Seleção de genitores baseada no comportamento em teste clonal

Acurácia

$$r_{aa_5} = \left[\frac{m N h_a^2}{1+(m-1) \rho + (N-1) m h_g^2} \right]^{1/2}$$

Ganho Genético

$$G_{s_5} = k r_{aa_5} \sigma_a$$

(1.6) Seleção de cruzamentos biparentais, baseada no comportamento da progénie de irmãos germanos e visando ao plantio (por meio de sementes) comercial destas (reconstituídas a cada geração)

Neste caso, a acurácia deve ser calculada em separado para os efeitos aditivos (r_{aa_6}) e de dominância (r_{dd}) e, posteriormente, obtém-se a acurácia para os efeitos genotípicos (r_{gg_6}), procedendo-se da seguinte maneira:

$$r_{aa_6} = \left[\frac{m N 0,25 h_a^2}{1+(m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]]} \right]^{1/2};$$

$$r_{dd_6} = \left[\frac{m N 0,0625 (h_g^2 - h_a^2)}{1+(m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]]} \right]^{1/2};$$

$$r_{gg_6} = \left[r_{aa_6}^2 + r_{dd_6}^2 \right]^{1/2};$$

Ganho genético

$$\begin{aligned} G_{s_6} &= k \ r_{\hat{g}g_6} \ \sigma_g \\ &= k \ r_{\hat{g}g_6} \ (h_g/h_a) \sigma_a \end{aligned}$$

Alternativamente, o ganho genético é dado por:

$$\begin{aligned} G_s &= k \ r_{\hat{a}a} \ \sigma_a + k \ r_{\hat{d}d} \ \sigma_d \\ &= k \left[r_{\hat{a}a} + r_{\hat{d}d} \left[\frac{h_g^2 - h_a^2}{h_a^2} \right]^{1/2} \right] \sigma_a \end{aligned}$$

- (1.7) Seleção de genitores baseada no efeito da capacidade geral de combinação sob cruzamentos fatoriais ou dialélicos

As expressões para a acurácia e ganho genético foram apresentadas por Resende (1999b) e encontram-se no Capítulo 15.

- (1.8) Seleção baseada em valores individuais e nas informações das famílias de irmãos germanos e de meios-irmãos obtidas sob cruzamentos fatoriais ou dialélicos

As expressões para a acurácia e ganho genético foram apresentadas por Resende (1999b) e encontram-se no Capítulo 15.

- (1.9) Idem (1.6), porém com seleção baseada em $(1/2) a_i + (1/2) a_j + d_{ij}$, em que a_i e d_{ij} referem-se à capacidade geral e específica de combinação, respectivamente, obtidas sob cruzamentos fatoriais ou dialélicos.

As expressões para a acurácia e ganho genético foram apresentadas por Resende (1999b) e encontram-se no Capítulo 15.

(2) Propagação assexuada dos indivíduos selecionados

- (2.1) Seleção baseada em valores individuais, sem o uso da informação da família ou de testes clonais

Acurácia	Ganho genético
$r_{\hat{g}g_7} = \left[\frac{m h_g^2}{1 + (m-1) \rho} \right]^{1/2}$	$\begin{aligned} G_{s_7} &= k \ r_{\hat{g}g_7} \ \sigma_g \\ &= k \ r_{\hat{g}g_7} \ (h_g/h_a) \sigma_a \end{aligned}$

(2.2) Seleção de indivíduos baseada no comportamento em teste clonal

Acurácia

$$r_{\hat{g}g_8} = \left[\frac{m N h_g^2}{1 + (m-1) \rho + (N-1) m h_a^2} \right]^{1/2}$$

Ganho genético

$$\begin{aligned} G_{s_8} &= k r_{\hat{g}g_8} \sigma_g \\ &= k r_{\hat{g}g_8} (h_g / h_a) \sigma_a \end{aligned}$$

(2.3) Seleção baseada em valores individuais e na informação da família de meios-irmãos

Acurácia

$$r_{\hat{g}g_9} = \left[\frac{m \frac{N-1}{N} [0,5625 h_a^2 + (h_g^2 - h_a^2)]}{1 + (m-1) \rho - m 0,25 h_a^2} + \frac{m [(1+(N-1) 0,25)^2 h_a^2 + (h_g^2 - h_a^2)]}{N [1 + (m-1) \rho + (N-1) m 0,25 h_a^2]} \right]^{1/2}$$

Ganho genético

$$\begin{aligned} G_{s_9} &= k r_{\hat{g}g_9} \sigma_g \\ &= k r_{\hat{g}g_9} (h_g / h_a) \sigma_a \end{aligned}$$

(2.4) Seleção baseada em valores individuais e na informação da família de irmãos germanos

Acurácia

$$r_{\hat{g}g_{10}} = \left[\frac{m \frac{N-1}{N} [0,25 h_a^2 + 0,5625 (h_g^2 - h_a^2)]}{1 + (m-1) \rho - m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]} + \frac{m [(1+(N-1) 0,5)^2 h_a^2 + (1+(N-1) 0,25)^2 (h_g^2 - h_a^2)]}{N [1 + (m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]]} \right]^{1/2}$$

Ganho genético

$$\begin{aligned} G_{s_{10}} &= k r_{\hat{g}g_{10}} \sigma_g \\ &= k r_{\hat{g}g_{10}} (h_g / h_a) \sigma_a \end{aligned}$$

(2.5) Seleção de indivíduos (clones) baseada no comportamento da progênie de meios-irmãos

Acurácia

$$r_{\hat{g}g_{11}} = 0,5 \left[\frac{m N h_a^2}{1 + (m-1) \rho + (N-1) m 0,25 h_a^2} \right]^{1/2}$$

Ganho genético

$$G_{s_{11}} = k r_{\hat{g}g_{11}} \sigma_a$$

(2.6) Seleção baseada em valores individuais e nas informações das famílias de irmãos germanos e de meios-irmãos obtidas sob cruzamentos fatoriais ou dialélicos

As expressões para a acurácia e ganho genético foram apresentadas por Resende (1999b) e encontram-se no Capítulo 15.

(3) Manutenção dos indivíduos selecionados no mesmo microambiente

(3.1) Seleção pelo valor fenotípico permanente

Acurácia

$$r_{\hat{p}p_{12}} = \left[\frac{m \rho}{1 + (m-1) \rho} \right]^{1/2}$$

Ganho com seleção

$$\begin{aligned} G_{s_{12}} &= k r_{\hat{p}p_{12}} \sigma_{fp} \\ &= k r_{\hat{p}p_{12}} (\rho/h_a^2)^{1/2} \sigma_a \end{aligned}$$

Nas expressões apresentadas, $N = nb$ refere-se ao número total de indivíduos por progênie, $r_{\hat{p}p}$ diz respeito à acurácia na predição dos efeitos fenotípicos permanentes e σ_{fp}^2 à variância destes efeitos.

As eficiências (por ciclo de seleção) do uso de medidas repetidas para um caráter com $\rho = 0,65$, considerando $N = 60$ para progênies e 20 para clones são apresentadas na Tabela 6.6 para alguns valores de herdabilidade e sistemas de seleção-propagação.

Tabela 6.6. Eficiência da utilização de 2 (EC_2), 3 (EC_3) e 4 (EC_4) medições em cada indivíduo em relação à utilização de apenas uma, visando ao melhoramento de um caráter quantitativo com repetibilidade de 0,65 e com herdabilidades no sentido restrito (h_a^2) e amplo (h_g^2) variáveis.

Sistema	h_a^2	h_g^2	EC2	EC3	EC4
1.1, 2.1 e 3.1	-	-	1,10	1,14	1,16
1.2	0,10	-	1,04	1,05	1,06
	0,20	-	1,02	1,03	1,04
	0,30	-	1,02	1,02	1,03
	0,40	-	1,01	1,02	1,02
2.2	-	0,15	1,02	1,03	1,04
	-	0,30	1,01	1,02	1,02
	-	0,45	1,01	1,01	1,01
	-	0,60	1,01	1,01	1,01
1.3	0,10	-	1,05	1,08	1,09
	0,20	-	1,05	1,08	1,09
	0,30	-	1,06	1,08	1,10
	0,40	-	1,07	1,09	1,11
2.3	0,10	0,15	1,06	1,09	1,10
	0,20	0,30	1,07	1,09	1,11
	0,30	0,45	1,07	1,10	1,12
	0,40	0,60	1,08	1,11	1,13
1.4	0,10	0,15	1,03	1,04	1,04
	0,20	0,30	1,03	1,04	1,04
	0,30	0,45	1,03	1,05	1,05
	0,40	0,60	1,04	1,06	1,07
2.4	0,10	0,15	1,03	1,04	1,05
	0,20	0,30	1,04	1,05	1,06
	0,30	0,45	1,05	1,07	1,08
	0,40	0,60	1,06	1,09	1,10

Constata-se que o ganho em eficiência com a utilização de várias medições depende do método de seleção e da herdabilidade do caráter, mesmo com um valor fixo e conhecido da repetibilidade. A eficiência é máxima para os métodos que utilizam apenas a informação do indivíduo e decresce rapidamente para os métodos que enfatizam mais a informação de médias de família ou de clones. Assim, o uso de duas medições propicia 10% a mais (em relação a uma medição) de ganho genético para as modalidades de seleção massal e, praticamente, não conduz a maior eficiência para a seleção de genitores (sistema 1.2) e de clones em testes clonais (sistema 2.2). Para os métodos que usam a informação do indivíduo e da família (sistemas 1.3, 2.3, 1.4 e 2.4), o ganho em eficiência reduz-se à metade (5%) ou menos.

A eficiência diminui com o aumento da herdabilidade para os métodos (1.2 e 2.2) que usam apenas as informações de médias e cresce com o aumento da

herdabilidade para os métodos (1.3, 2.3, 1.4 e 2.4) que usam tanto a informação de médias quanto a informação de indivíduos. Neste último caso, quanto maior o h_a^2 (*e* h_g^2) maior o peso dado à informação de indivíduo e portanto maior a eficiência. A mesma consideração é válida na comparação entre famílias de meios-irmãos e famílias de irmãos germanos, sendo que a eficiência é maior nas primeiras devido ao maior peso dado à informação do indivíduo no índice de seleção. Na seleção a partir de cruzamentos fatoriais a eficiência deverá ser inferior à eficiência na seleção com famílias de irmãos germanos visto que maior ênfase é dada nas informações de famílias de meios-irmãos e irmãos germanos obtidas sob cruzamentos fatoriais.

Os resultados revelam a necessidade de considerar tanto o método de seleção quanto as herdabilidades dos caracteres na recomendação do número adequado de medições em cada indivíduo e não apenas a repetibilidade. Os resultados apresentados nos tópicos anteriores ressaltam que é essencial o conhecimento dos parâmetros h_a^2 , h_g^2 e ρ (todos ao nível de indivíduos) associados aos caracteres sob melhoramento.

6.5. Repetibilidade de Caracteres com Distribuição Binomial

Para caracteres binários e tendo-se duas avaliações por indivíduo, uma tabela de contingência 2 x 2 sumariza toda a informação amostral. Tal tabela é estruturada da seguinte forma:

Avaliação 1

		0	1	Total
		n ₀₀	n ₁₀	n. ₀
Avaliação 2	0			
	1	n ₀₁	n ₁₁	n. ₁
Total		n _{..}	n _{1..}	n..

Com base na tabela de contingência apresentada, Lush (1956) relatou a estatística $\rho_L = n_{11}/n_1 - n_{01}/n_0$ como o estimador para a repetibilidade na escala binomial ou correlação entre avaliações repetidas em um mesmo indivíduo.

Para o caso da variável sobrevivência, podem ser especificadas as quantidades:

n₁₁ – número de indivíduos que estavam vivos na primeira avaliação e permaneceram vivos na segunda avaliação, ou seja, é o próprio número de indivíduos vivos na segunda avaliação.

n_{1..} – soma do número de indivíduos vivos na primeira avaliação e que morreram na segunda avaliação mais o número de indivíduos vivos na 2^a avaliação.

ação, ou seja, é o próprio número de indivíduos vivos na primeira avaliação.

n_{01} – número de indivíduos que estavam mortos na primeira avaliação e que se tornaram vivos na segunda avaliação, ou seja, zero.

Assim, neste caso, o estimador para a repetibilidade equivale a $\rho_L = N_2/N_1 = p_2/p_1$, em que N_2 e N_1 são os números de indivíduos vivos na segunda e primeira avaliações, respectivamente e p_2 e p_1 são as porcentagens de sobrevivência na segunda e primeira avaliações, respectivamente.

Como exemplo, considere que uma população experimental de dendezeiro tenha apresentado sobrevivências de 80% aos dois anos e de 70% aos três anos. Assim, a repetibilidade da sobrevivência nesta população equivale a $\rho_L = 0,70/0,80 = 0,875$.

6.6. Estimação da Repetibilidade pelo Método de Quadrados Mínimos

6.6.1. Amostra de dados não associada à estrutura experimental

O esquema de análise e os estimadores seguem o Modelo 3, apresentado no tópico 6.6.2, quando números iguais de medições são realizados em cada indivíduo.

Como exemplo, considere os resultados apresentados a seguir, associados à avaliação de 270 plantas com três medições cada uma.

FV	GL	QM	E(QM)
Medições (M)	2	$Q_1 = 338,2644$	$\sigma_t^2 + npbV_m$
Arvores (A)	269	$Q_2 = 27,8438$	$\sigma_t^2 + m\sigma_{ap}^2$
M x A	538	$Q_3 = 3,7399$	σ_t^2

Assim, têm-se as seguintes estimativas:

$$\hat{\sigma}_{ap}^2 = \frac{Q_2 - Q_3}{m} = \frac{27,8438 - 3,7339}{3} = 8,0346 : \text{variância permanente entre árvores.}$$

$$\hat{\sigma}_t^2 = Q_3 = 3,7399: \text{variância de ambiente temporário.}$$

$$\hat{\rho} = \frac{\sigma_{ap}^2}{\sigma_{ap}^2 + \sigma_t^2} = \frac{8,0346}{8,0346 + 3,7399} = 0,6824 : \text{repetibilidade individual.}$$

Para o caso de números desiguais de medições, o esquema de análise e os estimadores seguem o Modelo 4, apresentado no tópico 6.6.2. Como exem-

plo, considere os resultados apresentados a seguir, em que $m^* = 3$ refere-se à média harmônica do número de medições por indivíduo.

FV	GL	QM	E(QM)
Árvores	269	$Q_4 = 27,8438$	$\sigma_{t^*}^2 + m^* \sigma_{ap}^2$
Resíduo	540	$Q_5 = 4,9789$	$\sigma_{t^*}^2$

São encontradas as seguintes estimativas:

$$\hat{\sigma}_{ap}^2 = \frac{Q_4 - Q_5}{m^*} = \frac{27,8438 - 4,9789}{3} = 7,6216$$

$$\hat{\sigma}_{t^*}^2 = Q_5 = 4,9789$$

$$\hat{\rho} = \frac{\sigma_{ap}^2}{\sigma_{ap}^2 + \sigma_{t^*}^2} = \frac{7,6216}{7,6216 + 4,9789} = 0,6049$$

Ainda para o caso de dados não associados à estrutura experimental, Cruz & Regazzi (1994) apresentam, com detalhes, procedimentos de estimação de repetibilidade pelos métodos de componentes principais e da análise estrutural.

6.6.2. Dados associados ao delineamento experimental de blocos ao acaso

Estatisticamente, o coeficiente de repetibilidade equivale ao coeficiente de correlação intraclasse, o qual foi definido inicialmente por Harris (1913) e, posteriormente, recebeu um tratamento mais detalhado por Fisher (1948). Apesar de este tema ser abordado em vários livros de estatística (Gill, 1988; Barbin, 1993; Snedecor & Cochran, 1967; Steel & Torrie, 1980) e genética (Kempthorne, 1957; Turner & Young, 1969; Bulmer, 1980; Cardellino & Rovira, 1985; Pagnaco, 1997; Cruz & Regazzi, 1994), o coeficiente de correlação intraclasse individual associado a medidas repetidas e ao delineamento experimental de blocos ao acaso parece ter sido pouco estudado. Várias estratégias de análise são apresentadas a seguir, com base em Resende (1999b).

Modelo 1

O modelo estatístico associado à avaliação de p progênies em b blocos com n indivíduos por parcela e m medições por indivíduo equivale a:

$$Y_{ijk\ell} = \mu + p_i + b_j + e_{ij} + \delta_{jk} + m_\ell + t_{ijk\ell}, \text{ em que:}$$

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$.

p_i : efeito da progénie ou família i , aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$.

b_j : efeito do bloco j , fixo, $E(b_j) = b_j$ e $E(b_j^2) = V_b$.

e_{ij} : efeito da parcela ij , aleatório, $E(e_{ij}) = 0$ e $E(e_{ij}^2) = \sigma_e^2$.

δ_{ijk} : efeito do indivíduo k dentro da parcela ij , aleatório, $E(\delta_{ijk}) = 0$ e $E(\delta_{ijk}^2) = \sigma_\delta^2$.

m_ℓ : efeito da medição ℓ , fixo, $E(m_\ell) = 0$ e $E(m_\ell^2) = V_m$.

t_{ijkl} : efeito de ambiente temporário (erro), aleatório, $E(t_{ijkl}) = 0$ e $E(t_{ijkl}^2) = \sigma_t^2$.

O esquema de análise de variância associada ao modelo, bem como as esperanças matemáticas de quadrados médios ($E(QM)$) obtidas empregando-se o procedimento de Bennett & Franklin (1963) são apresentados a seguir, para a análise ao nível de medições individuais.

Fonte de Variação	GL	QM	E(QM)
Bloco	b-1	-	$\sigma_t^2 + m\sigma_\delta^2 + mn\sigma_e^2 + mnpV_b$
Progênie	p-1	Q_1	$\sigma_t^2 + m\sigma_\delta^2 + mn\sigma_e^2 + mnb\sigma_p^2$
Resíduo	(p-1)(b-1)	Q_2	$\sigma_t^2 + m\sigma_\delta^2 + mn\sigma_e^2$
Indivíduo (I)/Parcela	(n-1) pb	Q_3	$\sigma_t^2 + m\sigma_\delta^2$
Medição (M)	(m-1)	-	$\sigma_t^2 + npbV_m$
(I x M)	(m-1)(npb-1)	Q_4	σ_t^2
Total	mnpb - 1		

Para a análise ao nível de médias de parcela, o esquema é apresentado a seguir:

Fonte de Variação	GL	QM	E(QM)
Bloco	b-1	-	$\frac{\sigma_t^2}{mn} + \frac{\sigma_\delta^2}{n} + \sigma_e^2 + pV_b$
Progênie	p-1	Q_5	$\frac{\sigma_t^2}{mn} + \frac{\sigma_\delta^2}{n} + \sigma_e^2 + b\sigma_p^2$
Resíduo	(p-1)(b-1)	Q_6	$\frac{\sigma_t^2}{mn} + \frac{\sigma_\delta^2}{n} + \sigma_e^2$
Indivíduo (I)/Parcela	(n-1) pb	Q_7	$\sigma_t^2 + m\sigma_\delta^2$ ou $\frac{\sigma_t^2}{m} + \sigma_\delta^2$ (nível de médias de indivíduo)
Medição (M)	(m-1)	-	$\sigma_t^2 + npbV_m$
(I x M)	(m-1)(npb-1)	Q_8	σ_t^2
Total	mnpb - 1		

Considerando a análise mais comum, ao nível de médias de parcela, são os seguintes os estimadores para os componentes de variância:

$$\hat{\sigma}_t^2 = Q_8 : \text{variância de ambiente temporário.}$$

$$\hat{\sigma}_{\delta}^2 = (Q_7 - Q_8)/m : \text{variância permanente dentro de parcela.}$$

$$\hat{\sigma}_e^2 = Q_6 - Q_7/mn : \text{variância permanente entre parcelas.}$$

$$\hat{\sigma}_p^2 = (Q_5 - Q_6)/b : \text{variância genética entre progênies.}$$

$$\hat{\sigma}_y^2 = \sigma_p^2 + \sigma_e^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_t^2 : \text{variância fenotípica individual.}$$

E, ainda os seguintes estimadores:

$$\hat{h}_a^2 = 4\hat{\sigma}_p^2 / \hat{\sigma}_y^2 : \text{herdabilidade no sentido restrito (considerando progênies de meios-irmãos).}$$

$$\hat{\rho}_b = (\hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_{\delta}^2) / \hat{\sigma}_y^2 = (\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2 + \hat{\sigma}_e^2) / \hat{\sigma}_y^2 : \text{repetibilidade ao nível de indivíduo no bloco.}$$

$$\hat{\rho}_{\delta} = \hat{\sigma}_{\delta}^2 / (\hat{\sigma}_{\delta}^2 + \hat{\sigma}_t^2) = (3/4 \hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2) / (\hat{\sigma}_{\delta}^2 + \hat{\sigma}_t^2) : \text{repetibilidade ao nível de indivíduo na parcela.}$$

$$\hat{c}^2 = \hat{\sigma}_e^2 / \hat{\sigma}_y^2 : \text{correlação intraclasse devida ao ambiente permanente comum da parcela.}$$

$$\hat{\rho}_{mb} = \frac{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_{\delta}^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_t^2 / m} = \frac{m\hat{\rho}_b}{1 + (m-1)\hat{\rho}_b} : \text{repetibilidade ao nível de médias de safras, no bloco.}$$

$$\hat{\rho}_{m\delta} = \frac{\hat{\sigma}_{\delta}^2}{\hat{\sigma}_{\delta}^2 + \hat{\sigma}_{et}^2 / m} = \frac{m\hat{\rho}_{\delta}}{1 + (m-1)\hat{\rho}_{\delta}} : \text{repetibilidade ao nível de médias de safras, na parcela.}$$

$$\hat{h}_{in}^2 = \frac{m\hat{h}^2}{1 + (m-1)\hat{\rho}_b} : \text{herdabilidade ao nível de médias de indivíduo no bloco.}$$

Outro estimador importante para estudos referentes à determinação do tamanho ótimo de parcelas (Pimentel Gomes, 1984) é:

$$\hat{\rho}^* = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_e^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_t^2} : \text{correlação intraclasse (ou relação variância entre parcelas/(variância entre + variância dentro de parcelas)) devida ao ambiente permanente comum da parcela, considerando o efeito de progênies (tratamentos) como fixo.}$$

Os demais componentes de variância relatados apresentam os seguintes significados:

σ_a^2 , σ_d^2 e σ_{ep}^2 : variâncias genética aditiva, de dominância e ambiental permanente, respectivamente.

É importante notar que a variância de ambiente permanente ao nível de indivíduo (σ_{ep}^2) faz parte de σ_δ^2 .

Modelo 2

Para números desiguais de medições nos indivíduos, o seguinte modelo estatístico pode ser utilizado (Resende, 1999b):

$$Y_{ijk\ell} = \mu + p_i + b_j + e_{ij} + \delta_{ijk} + t_{ijk\ell}^*, \text{ em que:}$$

$t_{ijk\ell}^*$: efeito de ambiente temporário, aleatório, $E(t_{ijk\ell}^*)=0$ e $E(t_{ijk\ell}^{2*})=\sigma_{t^*}^2$

O esquema de análise de variância associada ao modelo é apresentado a seguir:

Fonte de Variação	GL	QM	E(QM)
Bloco	b-1	-	$\sigma_{t^*}^2 + m\sigma_\delta^2 + mn\sigma_e^2 + mnpV_b$
Progênie	p-1	Q ₉	$\sigma_{t^*}^2 + m\sigma_\delta^2 + mn\sigma_e^2 + mnbp\sigma_p^2$
Resíduo	(p-1)(b-1)	Q ₁₀	$\sigma_{t^*}^2 + m\sigma_\delta^2 + mn\sigma_e^2$
Indivíduo/parcela	(n-1) pb	Q ₁₁	$\sigma_{t^*}^2 + m\sigma_\delta^2$
Medições/Indivíduo	(m-1) npb	Q ₁₂	$\sigma_{t^*}^2$
Total	mnpb - 1		

Para a análise ao nível de médias de parcela, o esquema é apresentado a seguir:

Fonte de Variação	GL	QM	E(QM)
Bloco	b-1	-	$\frac{\sigma_{t^*}^2}{mn} + \frac{\sigma_\delta^2}{n} + \sigma_e^2 + pV_b$
Progênie	p-1	Q ₁₃	$\frac{\sigma_{t^*}^2}{mn} + \frac{\sigma_\delta^2}{n} + \sigma_e^2 + b\sigma_p^2$
Resíduo	(p-1)(b-1)	Q ₁₄	$\frac{\sigma_{t^*}^2}{mn} + \frac{\sigma_\delta^2}{n} + \sigma_e^2$
Indivíduo/parcela	(n-1) pb	Q ₁₅	$\sigma_{t^*}^2 + m\sigma_\delta^2$ ou $\frac{\sigma_{t^*}^2}{m} + \sigma_\delta^2$ (nível de médias)
Medições/indivíduo	(m-1) npb	Q ₁₆	$\sigma_{t^*}^2$
Total	mnpb - 1		

Estimadores similares aos obtidos para o modelo 1 podem ser obtidos, bastando substituir Q_5 , Q_6 , Q_7 e Q_8 por Q_{13} , Q_{14} , Q_{15} e Q_{16} , respectivamente, e trabalhando-se m como a média harmônica do número de medições por indivíduo.

Modelo 3

Ignorando-se a estrutura do delineamento experimental, tem-se o seguinte modelo (Turner & Young, 1969; Cruz & Regazzi, 1994) para o caso de números iguais de medições:

$$Y_{ij} = \mu + m_i + ap_j + t_{ij}, \text{ em que:}$$

m_i : efeito da medição i , fixa, $E(m_i) = m_i$ e $E(m_i^2) = V_m$.

ap_j : efeito permanente da árvore j , aleatório, $E(ap_j) = 0$ e $E(ap_j^2) = \sigma_{ap}^2$.

t_{ij} : efeito de ambiente temporário, aleatório, $E(t_{ij}) = 0$ e $E(t_{ij}^2) = \sigma_t^2$.

O esquema de análise de variância associada ao modelo é apresentado a seguir:

Fonte de Variação	GL	QM	E(QM)
Medição (M)	$m-1$	-	$\sigma_t^2 + npbV$
Árvores (A)	$npb-1$	Q_{17}	$\sigma_t^2 + m\sigma_{ap}^2$
(M x A)	$(m-1)(npb-1)$	Q_{18}	σ_t^2

Neste caso, os estimadores são:

$$\hat{\sigma}_{ap}^2 = \frac{Q_{17} - Q_{18}}{m}; \quad \hat{\sigma}_t^2 = Q_{18}$$

$$\hat{\rho}_e = \frac{\hat{\sigma}_{ap}^2}{\hat{\sigma}_{ap}^2 + \hat{\sigma}_t^2} : \text{repetibilidade ao nível de indivíduos no experimento.}$$

Observa-se, também, a igualdade $\sigma_{ap}^2 = \sigma_p^2 + \sigma_\delta^2 + \sigma_e^2 + \sigma_b^2$, de forma que ρ_e inclui também a variância entre blocos, como parte da variância fenotípica permanente e da variância fenotípica total. Este mesmo estimador pode ser obtido a partir do Modelo 1, considerando o efeito de bloco como aleatório.

O Modelo 1 é preferível em relação ao Modelo 3, pois permite estimar a repetibilidade ao nível de indivíduo no bloco, a qual é mais útil na prática por ser usada nos índices multiefeitos e no procedimento BLUP para a predição de valores genéticos.

Modelo 4

Para o caso de números desiguais de medições nos indivíduos e ignorando a estrutura experimental, tem-se o modelo estatístico (Kempthorne, 1957; Snedecor & Cochran, 1967; Vencovsky, 1972; Cruz & Regazzi, 1994):

$$Y_{ij} = \mu + ap_j + t_{ij}^*$$

O esquema de análise de variância associado ao modelo é apresentado a seguir:

Fonte de Variação	GL	QM	E(QM)
Árvores	npb-1	Q ₁₉	$\sigma_{t^*}^2 + m\sigma_{ap}^2$
Resíduo	(m-1) npb	Q ₂₀	$\sigma_{t^*}^2$

De maneira similar à descrita para o Modelo 3, obtém-se o estimador para a repetibilidade individual no experimento: $\hat{\rho}_e = \frac{\hat{\sigma}_{ap}^2}{\hat{\sigma}_{ap}^2 + \hat{\sigma}_{t^*}^2}$. Estimador idêntico pode ser obtido pelo Modelo 2, considerando o efeito de bloco como aleatório.

Modelo 5

Os modelos 1 e 2 consideram que as interações progênies x medições e medições x blocos não são significativas. O modelo fatorial completo para a análise ao nível indivíduos em uma determinada medição é dado por:

$$Y_{ijkl} = \mu + p_i + b_j + m_\ell + pb_{ij} + pm_{i\ell} + bm_{j\ell} + pbm_{j\ell} + \delta_{ijkl}^*, \text{ em que:}$$

pb_{ij} : efeito da interação progênies x blocos ou efeito e_{ij} da parcela ij, aleatório,

$$E(pb_{ij})=0 \text{ e } E(pb_{ij}^2)=\sigma_{pb}^2$$

$pm_{i\ell}$: efeito da interação progênies x medições, aleatório,

$$E(pm_{i\ell})=0 \text{ e } E(pm_{i\ell}^2)=\sigma_{pm}^2$$

$bm_{j\ell}$: efeito da interação blocos x medições, fixo,

$$E(bm_{j\ell})=bm_{j\ell} \text{ e } E(bm_{j\ell}^2)=V_{bm}$$

$pbm_{j\ell}$: efeito da interação progênies x blocos x medições, aleatório,

$$E(pbm_{j\ell})=0 \text{ e } E(pbm_{j\ell}^2)=\sigma_{pbm}^2$$

δ_{ijkl}^* : efeito do indivíduo k dentro da parcela ij, na medição ℓ , aleatório,

$$E(\delta_{ijkl}^*)=0 \text{ e } E(\delta_{ijkl}^{*2})=\sigma_{\delta^*}^2$$

O efeito δ_{ijkl}^* pode ser decomposto em $\delta_{ijkl}^* = \delta_{ijk} + t_{ijkl}^{**}$, em que:

δ_{ijk} : efeito do indivíduo K dentro da parcela ij, aleatório,
 $E(\delta_{ijk})=0$ e $E(\delta_{ijk}^2)=\sigma_\delta^2$

t_{ijkl}^{**} : efeito da interação indivíduos x medições, dentro de parcelas, aleatório,

$E(t_{ijkl}^{**})=0$ e $E(t_{ijkl}^{**2})=\sigma_{t^{**}}^2$

O esquema da análise de variância associada ao modelo fatorial completo ao nível de indivíduos é apresentado a seguir:

Fonte de Variação	GL	QM	E(QM)	F
Bloco (B)	(b-1)	Q ₂₈	$\sigma_{\delta^*}^2 + nm \frac{b}{b-1} \sigma_{pb}^2 + nmp V_b$	Q ₂₈ /Q ₃₀
Progênie (P)	(p-1)	Q ₂₉	$\sigma_{\delta^*}^2 + nbm \sigma_{p^*}^2$	Q ₂₉ /Q ₃₅
P x B	(p-1) (b-1)	Q ₃₀	$\sigma_{\delta^*}^2 + nm \frac{b}{b-1} \sigma_{pb}^2$	Q ₃₀ /Q ₃₅
Medição (M)	(m-1)	Q ₃₁	$\sigma_{\delta^*}^2 + nb \frac{m}{m-1} \sigma_{pm}^2 + npb V_m$	Q ₃₁ /Q ₃₂
P x M	(p-1) (m-1)	Q ₃₂	$\sigma_{\delta^*}^2 + nb \frac{m}{m-1} \sigma_{pm}^2$	Q ₃₂ /Q ₃₅
B x M	(b-1) (m-1)	Q ₃₃	$\sigma_{\delta^*}^2 + n \frac{b}{b-1} \frac{m}{m-1} \sigma_{pbm}^2 + np V_{bm}$	Q ₃₃ /Q ₃₄
P x B x M	(p-1)(b-1)(m-1)	Q ₃₄	$\sigma_{\delta^*}^2 + n \frac{b}{b-1} \frac{m}{m-1} \sigma_{pbm}^2$	Q ₃₄ /Q ₃₅
Indivíduo (I)/(PBM)	(n-1) pbm	Q ₃₅	$\sigma_{\delta^*}^2$	-
- Indivíduo (I)/Parcela	(n-1) pb	Q ₃₆	$\sigma_{t^{**}}^2 + m \sigma_\delta^2$	-
- (I x M)/ Parcela	(n-1)(m-1) pb	Q ₃₇	$\sigma_{t^{**}}^2$	-

Com base neste esquema de análise, são os seguintes os estimadores para os componentes de variância:

$$\hat{\sigma}_{t^{**}}^2 = Q_{37}; \quad \hat{\sigma}_{pbm}^2 = (Q_{34}-Q_{35})/(n \frac{b}{b-1} \frac{m}{m-1}); \quad \hat{\sigma}_{pb}^2 = (Q_{30}-Q_{35})/(nm \frac{b}{b-1});$$

$$\hat{\sigma}_d^2 = (Q_{36}-Q_{37})/m; \quad \hat{\sigma}_{pm}^2 = (Q_{32}-Q_{35})/(nb \frac{m}{m-1}); \quad \hat{\sigma}_{p^*}^2 = (Q_{29}-Q_{35})/(nbm);$$

Há ainda os seguintes estimadores, referentes a:

$$\hat{\sigma}_{et}^2 = \hat{\sigma}_{t^{**}}^2 + \hat{\sigma}_{pbm}^2 + \hat{\sigma}_{pm}^2 : \text{variância de ambiente temporário.}$$

$$\hat{\sigma}_{\delta}^2 : \text{variância permanente dentro de parcela.}$$

$$\hat{\sigma}_{pb}^2 : \text{variância permanente entre parcelas.}$$

$$\hat{\sigma}_{p^*}^2 : \text{variância entre progênies, inflacionada pelas interações progênies x medições e progênies x blocos.}$$

$$\hat{\sigma}_{p^{**}}^2 = \hat{\sigma}_{p^*}^2 - \hat{\sigma}_{pm}^2 - \hat{\sigma}_{pb}^2 : \text{variância entre progênies livres das interações com medições e com blocos.}$$

$$\hat{\sigma}_y^2 = \hat{\sigma}_{t^{**}}^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_{pbm}^2 + \hat{\sigma}_{pm}^2 + \hat{\sigma}_{pb}^2 + \hat{\sigma}_p^2 : \text{variância fenotípica.}$$

$$\hat{h}_a^2 = 4\hat{\sigma}_{p^{**}}^2 / \hat{\sigma}_y^2 : \text{herdabilidade individual no sentido restrito (considerando progênies de meios-irmãos).}$$

$$\hat{\rho} = (\hat{\sigma}_{p^{**}}^2 + \hat{\sigma}_{pb}^2 + \hat{\sigma}_{\delta}^2) / \hat{\sigma}_y^2 : \text{repetibilidade individual.}$$

$$\hat{c}^2 = \hat{\sigma}_{pb}^2 / \hat{\sigma}_y^2 : \text{correlação devida ao ambiente comum da parcela.}$$

$$\hat{p}^2 = (\hat{\sigma}_{\delta}^2 - 3\hat{\sigma}_{p^{**}}^2) / \hat{\sigma}_y^2; \quad \hat{pm}^2 = \hat{\sigma}_{pm}^2 / \hat{\sigma}_y^2.$$

As correlações genéticas através das medições e dos blocos são dadas por:

$$\hat{r}_{a_m} = \hat{\sigma}_{p^{**}}^2 / (\hat{\sigma}_{p^{**}}^2 + \hat{\sigma}_{pm}^2) : \text{correlação genética do caráter com ele mesmo, através das medições.}$$

$$\hat{r}_{a_b} = \hat{\sigma}_{p^{**}}^2 / (\hat{\sigma}_{p^{**}}^2 + \hat{\sigma}_{pb}^2) : \text{correlação genética do caráter com ele mesmo, através dos blocos.}$$

$$\hat{r}_{a_{mb}} = \hat{\sigma}_{p^{**}}^2 / (\hat{\sigma}_{p^{**}}^2 + \hat{\sigma}_{pm}^2 + \hat{\sigma}_{pb}^2 + \hat{\sigma}_{pbm}^2) : \text{correlação genética do caráter com ele mesmo, através das medições, dos blocos e das combinações medições-bloco.}$$

Os estimadores de correlações (r_a) apresentados representam o limite inferior da correlação genética esperada para um caráter avaliado em diferentes ambientes (medições e/ou blocos). Tais estimadores subestimam a correlação genética se houver heterogeneidade de variâncias. Com correção prévia dos dados para a heterogeneidade de variâncias, estes estimadores são exatos. Sem tal correção prévia, os estimadores são conservadores.

É importante relatar que o próprio uso do modelo de repetibilidade pressupõe que a correlação genética do caráter com ele mesmo em outra medição equivale a 1. Neste caso, tal modelo admite que a interação genótipo x medições pode ser devido apenas à heterogeneidade de variâncias. Assim sendo, os modelos 1 e 2 que ignoram a interação progênies x medições são coerentes. Quando a correlação genética de um caráter com ele mesmo em outra medição

for menor que 1, o uso de um modelo multivariado (que assume diferentes medições como diferentes caracteres) é recomendado. Os mesmos comentários são válidos para correlação genética do caráter com ele mesmo em outro bloco.

Como exemplo, considere os quadrados médios associados à avaliação (para o caráter produção de massa foliar) de $p = 15$ famílias de meios irmãos de erva-mate, em $b = 6$ blocos casualizados, com 3 plantas por parcela e 3 medições por indivíduo. A partir destes resultados, são estimados, simultaneamente, a herdabilidade e a repetibilidade do caráter, considerando o modelo 1.

Tabela 6.7. Quadrados médios (QM) e esperanças (E) associados à análise de variância ao nível de médias de parcela envolvendo medidas repetidas em espécies perenes.

Fontes de variação	GL	QM	E(QM)
Bloco	(b-1) = 5	$Q_1 = 31,47208$	—
Família	(p-1) = 14	$Q_2 = 4,62181$	$\frac{\sigma_t^2}{nm} + \frac{\sigma_\delta^2}{n} + \sigma_e^2 + b\sigma_p^2$
Resíduo	(b-1)(p-1)=70	$Q_3 = 3,10120$	$\frac{\sigma_t^2}{nm} + \frac{\sigma_\delta^2}{n} + \sigma_e^2$
Indivíduo (I)/Parcela	(n-1) pb = 180	$Q_4 = 19,65308$	$\frac{\sigma_t^2}{m} + \sigma_\delta^2$
Medição (M)	(m-1) = 2	$Q_5 = 338,26439$	—
I x M	(m-1)(npb-1) = 538	$Q_6 = 3,73997$	σ_t^2
Total	mnpb-1 = 809		

A obtenção dos quadrados médios (QM) pode ser realizada em três etapas: (i) análise de variância ao nível de médias de parcela (Vencovsky & Barriaga, 1992) para obtenção dos QM's relativos às três primeiras fontes de variação; (ii) obtenção do QM Indivíduo (I)/Parcela trabalhando-se com médias de medições de indivíduo; (iii) análise de variância de dados individuais em classificação cruzada ignorando a estrutura experimental (Cruz & Regazzi, 1994) e a fonte de variação indivíduo (esta deve ser computada e depois ignorada), para obtenção dos dois últimos QM's da Tabela 6.7.

Com base na Tabela 6.7, as seguintes estimativas de componentes de variância e parâmetros genéticos podem ser obtidas:

$$\hat{\sigma}_t^2 = Q_6 = 3,73997: \text{variância ambiental temporária.}$$

$$\hat{\sigma}_\delta^2 = (3/4) \hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2 = Q_4 - Q_6/m = 5,30437: \text{variância permanente dentro de parcelas.}$$

$\hat{\sigma}_e^2 = Q_3 - Q_4/n = 0,91753$: variância permanente entre parcelas ou variância do ambiente comum da parcela.

$\hat{\sigma}_p^2 = (1/4) \hat{\sigma}_a^2 = (Q_2 - Q_3)/b = 0,25349$: variância genética entre famílias de meios-irmãos.

$\hat{h}_a^2 = \frac{4\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_t^2} = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_t^2} = 0,0992$: herdabilidade no sentido restrito.

$\hat{\rho} = \frac{\hat{\sigma}_p^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_e^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_t^2} = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2 + \hat{\sigma}_e^2}{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{ep}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_t^2} = 0,6339$: repetibilidade individual.

$\hat{c}^2 = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{\delta}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_t^2} = 0,0896$: correlação intraclasse devida ao ambiente comum da parcela.

A herdabilidade no sentido amplo pode ser estimada a partir de teste clonal segundo o mesmo delineamento e procedimentos de análises similares aos descritos para a estimativa da herdabilidade no sentido restrito.

6.7. Estimação da Repetibilidade pelo Método de Máxima Verossimilhança Restrita (REML)

O método (Patterson e Thompson, 1971; Searle et al., 1992) da máxima verossimilhança restrita (REML) é, atualmente, o método padrão para a estimativa de componentes de variância a partir de dados desbalanceados em espécies perenes. A seguir são apresentados os estimadores para os componentes de variância associados aos modelos apresentados no tópico 6.6.2.

Em termos matriciais, o modelo (5) equivale a:

$$y = Xb + Za + Wc + Tp + Qs + Ru + e, \text{ em que:}$$

y, b, a, c, p, s, u e e = vetores de dados, de efeitos fixos (bloco, medição e interação bloco x medição, ajustados em um único efeito denominado combinação bloco-medição com mb elementos e $mb-1$ graus de liberdade), de valores adicionais, de parcela (interação família x bloco), de efeitos permanentes, da interação família x medição (ou safra), da interação família x bloco x medição e de erros aleatórios (ambiente temporário), respectivamente.

X, Z, W, T, Q e R : matrizes de incidência para b, a, c, p, s e u , respectivamente.

As equações de modelo misto, para a predição BLUP dos efeitos aleatórios, equivalem a:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \\ \hat{s} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q & X'R \\ Z'X & ZZ + A^{-1}\lambda_1 & ZW & ZT & ZQ & ZR \\ W'X & W'Z & WW + I\lambda_2 & WT & WQ & WR \\ T'X & TZ & TW & TT + I\lambda_3 & TQ & TR \\ Q'X & Q'Z & Q'W & QT & QQ + I\lambda_4 & QR \\ R'X & R'Z & RW & RT & RQ & RR + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ Q'y \\ R'y \end{bmatrix}$$

em que:

A : matriz de correlação genética aditiva entre parentes.

I : matriz identidade, de ordem apropriada.

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_c^2}; \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_p^2}; \quad \lambda_4 = \frac{\sigma_e^2}{\sigma_s^2}; \quad \lambda_5 = \frac{\sigma_e^2}{\sigma_u^2}.$$

$\sigma_a^2 = 4\sigma_p^2$: variância genética aditiva.

$\sigma_c^2 = \sigma_{pb}^2$: variância entre parcelas.

$\sigma_p^2 = \sigma_{ep}^2$: variância de ambiente permanente.

$\sigma_s^2 = \sigma_{pm}^2$: variância da interação progênies x medição.

$\sigma_u^2 = \sigma_{pmb}^2$: variância da interação progênie x medição x bloco.

σ_e^2 : variância de ambiente temporário.

Os estimadores dos componentes de variância pelo método REML (algoritmo EM) equivalem a:

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

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

$$\hat{\sigma}_c^2 = [\hat{c}'c + \hat{\sigma}_e^2 \operatorname{tr}(C^{cc})] / (pb)$$

$$\hat{\sigma}_p^2 = [(\hat{p}'p) + \hat{\sigma}_e^2 \operatorname{tr}(C^{pp})] / (npb)$$

$$\hat{\sigma}_s^2 = [(\hat{s}'s) + \hat{\sigma}_e^2 \operatorname{tr}(C^{ss})] / (pm)$$

$$\hat{\sigma}_u^2 = [(\hat{u}'u) + \hat{\sigma}_e^2 \operatorname{tr}(C^{uu})] / (pbm)$$

$r(X)$: posto da matriz X .

tr : operador traço matricial.

$C^{aa}, C^{cc}, C^{pp}, C^{ss}$ e C^{uu} : partição da inversa da matriz dos coeficientes das equações de modelo misto, referentes aos efeitos a, c, p, s e u, respectivamente.

Na análise pelo software DFREML (Meyer, 1998), o arquivo deve apresentar a seguinte seqüência de colunas:

Indivíduo Pai Mãe Bloco Medição B x M Parcela Permanente PxM PxMxB Dados

Para o Modelo (1) é o seguinte o modelo linear misto:

$$y = X_1 b^* + Za + Wc + Tp + e^*, \text{ em que:}$$

b^* : vetor de efeitos fixos, o qual contempla apenas os efeitos de blocos e de medições.

e^* : vetor de erros, o qual contempla também os efeitos (BxM), s e u.

X_1 : matriz de incidência para b^* .

As equações de modelo misto e os estimadores dos componentes de variância seguem o sistema matricial descrito para o Modelo 5, bastando ignorar as duas últimas linhas e duas últimas colunas e substituir b por b^* e σ_e^2 por $\sigma_{e^*}^2$ nos estimadores e preditores.

O arquivo para análise do Modelo 1 no software DFREML deve apresentar a seguinte seqüência de colunas:

Indivíduo Pai Mãe Bloco Medição Parcela Permanente Dados

Um aperfeiçoamento em relação ao Modelo 1 pode ser conseguido ajustando-se a combinação bloco-medição no vetor de efeitos fixos. Neste caso, as previsões são realizadas após o ajuste também para o efeito (BxM), o qual não será, então, incluído no vetor de erros.

Para o Modelo 2, o modelo linear associado equivale a:

$$y = X_2 b^{**} + Za + Wc + Tp + e^{**}, \text{ em que:}$$

b^{**} : vetor de efeitos fixos, o qual contempla apenas o efeito de bloco.

e^{**} : vetor de erros, o qual contempla $e^* +$ o efeito de medição.

X_2 : matriz de incidência para b^{**} .

As equações de modelo misto e os estimadores dos componentes de variância seguem conforme descrito para o Modelo 1, bastando substituir b^* e $\sigma_{e^*}^2$ por b^{**} e $\sigma_{e^{**}}^2$, respectivamente.

O arquivo para análise do Modelo 2 no software DFREML deve apresentar a seguinte seqüência de colunas:

Indivíduo Pai Mãe Bloco Parcela Permanente Dados

O Modelo 3, em notação matricial é dado por:

$$y = X_3 b^{***} + Zp^* + e^*. \text{ em que:}$$

b^{***} : vetor de efeitos fixos, o qual contempla apenas o efeito de medição.

e^* : vetor aleatório de erros.

p^* : vetor de efeitos permanentes, o qual contempla os efeitos de ambiente permanente, dominância e aditivos (os efeitos de bloco e de parcela são também contemplados neste vetor, como efeitos ambientais permanentes).

X_3 : matriz de incidência para b^{***} .

As equações de modelo misto são:

$$\begin{bmatrix} \hat{b}^{***} \\ \hat{p}^* \end{bmatrix} = \begin{bmatrix} X'_3 & X'_3 Z \\ Z' X_3 & Z' Z + I \end{bmatrix}^{-1} \begin{bmatrix} X'_3 y \\ Z' y \end{bmatrix}, \text{ em que } \lambda_6 = \sigma_{e^{***}}^2 / \sigma_p^2$$

O arquivo para análise do Modelo 3 no software DFREML deve apresentar a seguinte seqüência de colunas:

Indivíduo Pai Mãe Medição Dados

6.8. Estimação Comparativa da Repetibilidade por Diferentes Modelos

Considere os resultados obtidos a partir da avaliação do caráter produção de massa foliar em um experimento balanceado de erva-mate no delineamento de blocos casualizados com 15 progênies, 6 blocos, 3 plantas por parcela e 3 medições por indivíduo.

A partir desses dados foram estimados os componentes de variância e os parâmetros genéticos e fenotípicos, empregando-se os softwares Selegen (Resende et al., 1994a) e DFREML (Meyer, 1998).

Nas Tabelas 6.8 e 6.9 são apresentados os quadrados médios associados à análise ao nível de médias de parcela e aos Modelos 1 e 2, respectivamente.

Tabela 6.8. Quadrados médios (QM) para a análise de variância ao nível de médias de parcela, associada ao Modelo 1.

Fonte de Variação	GL	QM
Bloco	5	31,47208
Progénie	14	4,62181
Resíduo	70	3,10120
Indivíduo (I) Parcela	180	19,65308
Medição (M)	2	338,26439
(I x M)	538	3,73997

Tabela 6.9. Quadrados médios (QM) para a análise de variância ao nível de médias de parcela, associada ao Modelo 2.

Fonte de Variação	GL	QM
Bloco	5	31,47208
Progénie	14	4,62181
Resíduo	70	3,10120
Indivíduo/Parcela	180	19,65308
Medições/Indivíduo	540	4,97895

Nas Tabelas 6.10 e 6.11 são apresentados os quadrados médios associados à análise de variância pelos Modelo 3 e 4 respectivamente.

Tabela 6.10. Quadrados médios (QM) para a análise de variância associada ao Modelo 3.

Fonte de Variação	GL	QM
Medições (M)	2	338,26439
Árvores (A)	269	27,84376
M x A	538	3,73997

Tabela 6.11. Quadrados médios (QM) para a análise de variância associada ao Modelo 4.

Fonte de Variação	GL	QM
Árvores	269	27,84376
Resíduo	540	4,978948

Verifica-se que os 4 modelos de análise apresentam quadrados médios equivalentes para algumas fontes de variação. Na Tabela 6.12 são apresentados os componentes de variância e parâmetros estimados com base no modelo 1 e também empregando-se o método de máxima verossimilhança restrita (REML).

Tabela 6.12. Estimativas de componentes de variância e parâmetros genéticos e fenotípicos pelo modelo 1 (software SELEGEN) e também por máxima verossimilhança restrita (REML) empregando-se o software DFREML (ajustando bloco e medição como efeitos fixos e parcela, efeitos permanentes e efeitos adicionais como aleatórios).

SELEGEN		DFREML	
Componentes de variância	Parâmetros	Parâmetros	Componentes de variância
$\hat{\sigma}_p^2 = 0,25349$	$\hat{\rho}_b = 0,6339$	$\rho_b = 0,6338$	$\hat{\sigma}_y^2 = 10,21497$
$\hat{\sigma}_e^2 = 0,91753$	$\hat{h}_a^2 = 0,0992$	$\hat{n}_a^2 = 0,0998$	$\hat{\sigma}_a^2 = 1,01902$
$\hat{\sigma}_{\delta}^2 = 5,30437$	$\hat{c}^2 = 0,0896$	$\hat{c}^2 = 0,0898$	$\hat{\sigma}_e^2 = 0,91731$
$\hat{\sigma}_t^2 = 3,73997$			$\hat{\sigma}_t^2 = 3,74028$
$\hat{\sigma}_y^2 = 10,21531$			$\hat{\sigma}_{ep}^2 + \hat{\sigma}_d^2 = 4,53837$
$\hat{\sigma}_a^2 = 1,01376$			
$\hat{\sigma}_{ep}^2 + \hat{\sigma}_d^2 = 4,54405$			

Constata-se a equivalência entre os métodos de quadrados mínimos e REML, conforme esperado para o caso balanceado (Tabela 6.12). Na Tabela 6.13 são apresentados os componentes de variância e parâmetros estimados com base no Modelo 2 e também empregando-se o método REML.

Tabela 6.13. Estimativas de componentes de variância e parâmetros genéticos e fenotípicos pelo modelo 2 (software SELEGEN) e também por máxima verossimilhança restrita (REML) empregando-se o software DFREML (ajustando bloco como efeito fixo e parcela, efeitos permanentes e efeitos aditivos como aleatórios).

SELEGEN		DFREML
Componentes de variância	Parâmetros	Parâmetros
$\hat{\sigma}_p^2 = 0,25344$	$\hat{\rho}_b = 0,5491$	$\hat{\rho}_b = 0,5491$
$\hat{\sigma}_e^2 = 0,91753$	$\hat{h}_a^2 = 0,0918$	$\hat{h}_a^2 = 0,0925$
$\hat{\sigma}_{\delta}^2 = 4,89138$	$\hat{c}^2 = 0,0831$	$\hat{c}^2 = 0,0829$
$\hat{\sigma}_{t*}^2 = 4,97895$		$\hat{\sigma}_y^2 = 11,04225$
$\hat{\sigma}_y^2 = 11,0413$		

Verifica-se que o modelo 1 conduz a uma repetibilidade de maior magnitude, pois não inclui a variância entre medições em $\hat{\sigma}_t^2$, fato que o faz o Modelo 2 (Tabelas 6.12.e.6.13). Na Tabela 6.14 são apresentados os resultados obtidos pelos Modelos 3 e 4.

Tabela 6.14. Estimativas de componentes de variância e da repetibilidade pelos Modelos 3 e 4 (software Selegen) e empregando-se o software DFREML (ajustando o efeito de medição como fixo e o efeito permanente como aleatório).

Modelo 3			Modelo 4	
Componentes de variância	Repetibilidade	DFREML	Componentes de variância	Repetibilidade
$\hat{\sigma}_{ap}^2 = 8,03460$	$\hat{\rho}_e = 0,6824$	$\hat{\rho}_e = 0,6825$	$\hat{\sigma}_{ap}^2 = 7,62160$	$\hat{\rho}_e = 0,6049$
$\hat{\sigma}_t^2 = 3,73997$		$\hat{\sigma}_y^2 = 11,78871$	$\hat{\sigma}_{t*}^2 = 4,97895$	
$\hat{\sigma}_y^2 = 11,77457$			$\hat{\sigma}_y^2 = 12,60055$	

De maneira geral, verifica-se que os Modelos 3 e 4 superestimam a repetibilidade em relação aos Modelos 1 e 2, respectivamente. Assim, os Modelos 1 e 2 devem ser preferencialmente utilizados.

O Modelo 1 de análise deve ser empregado de forma rotineira no melhoramento de espécies perenes, pois nestas espécies são efetuadas várias medições em cada planta do experimento, e, raramente, os experimentos são desativados após uma única avaliação. Ademais este modelo permite a estimativa simultânea da herdabilidade e da repetibilidade.

Os resultados referentes aos modelos fatoriais completo (5) e parciais são apresentados na Tabela 6.15.

Tabela 6.15. Estimativas de componentes de variância e parâmetros genéticos pelos modelos fatoriais completo e parciais, empregando-se o software DFREML.

Parâmetros	Modelo 1	Modelo F ₁ *	Modelo F ₂ **	Modelo F ₃ ***
\hat{h}_a^2	0,0992	0,0991	0,0992	0,0994
$\hat{\rho}$	0,6339	0,6342	0,6742	0,6740
\hat{c}^2	0,0898	0,0895	0,0900	0,0906
\hat{p}^2	0,4444	0,4456	0,4850	0,4840
\hat{pm}^2	-	0,0016	0,0080	-
Log L	-1.186,250	-1.185,020	-1.146,725	-1.147,725

* Modelo F₁ = modelo ajustando as interações progênies x blocos e progênies x medições; ** Modelo F₂ = modelo ajustando as 3 interações; *** Modelo F₃ = modelo ajustando as interações progênies x blocos e blocos x medições.

Verifica-se que o componente de variância da interação progênies x medições é desprezível, representando menos de 1% (pm^2) da variação fenotípica total, sendo que as correlações r_{am} equivalem a 0,94 e 0,76 para os modelos F₁ e F₂, respectivamente (Tabela 6.15).

O ajuste do efeito fixo da interação blocos x medições conduziu a um ligeiro aumento da estimativa do coeficiente de repetibilidade e melhorou o ajuste do modelo. As estimativas da herdabilidade praticamente não foram afetadas pelo ajuste ou não das interações progênies x medições e medições x blocos (Tabela 6.15).

É importante relatar que os resultados foram obtidos a partir de dados simulados. Com dados reais, na maioria das vezes, não é esperada a significância da interação blocos x medições dentro de um mesmo experimento e, portanto, o ajuste para esta interação não seria necessário. Sendo as condições edáficas dos blocos idênticas de uma medição para outra e as condições climáticas (precipitação por exemplo) iguais para todos os blocos em uma medição, não é esperado que determinados blocos sejam melhores em uma medição e outros blocos sejam os melhores em outras medições (o que caracterizaria uma interação). De fato, empregando-se dados reais de acerola, envolvendo o caráter pe-

so de frutos avaliados em 12 meses e três blocos, constatou-se a não significância da interação bloco x medição, conforme os resultados da Tabela 6.16 (Paiva et al., 2001).

Assim, desprezando a interação B x M e tendo-se em vista que os componentes σ_{pm}^2 , σ_{pbm}^2 e σ_t^2 dizem respeito a componentes da variância de ambiente temporário, e portanto, podem ser reunidos em um único resíduo σ_t^2 , tem-se que o modelo de análise proposto (Modelo 1) pode ser utilizado eficientemente. Também, outra opção, em vez de desprezar a interação B x M, refere-se ao uso do próprio Modelo 1, porém substituindo as fontes de variação referentes a bloco e medição por uma fonte de variação associada à combinação bloco - medição.

Tabela 6.16. Estimativas de parâmetros genéticos e fenotípicos associados à avaliação de 62 progênies de acerola, em três blocos casualizados com quatro plantas por parcela e 12 avaliações mensais por indivíduo. Caráter peso de frutos.

Parâmetros*	Modelo 1	Modelo incluindo a interação B x M
\hat{h}_a^2	0,081	0,080
\hat{c}^2	0,020	0,020
\hat{p}^2	0,073	0,076
$\hat{\rho}$	0,174	0,176

Repetibilidade ao nível de médias de safras

A análise completa dos dados de acerola, pelo Modelo 1, é apresentada na Tabela 6.17.

Tabela 6.17. Análise dos dados de produção de frutos em acerola.

Fonte de Variação	GL	QM	E(QM)
Bloco	2	-	-
Progénie	61	$Q_1 = 19520921,80$	$\sigma_t^2 + m\sigma_\delta^2 + mn\sigma_e^2 + mnb\sigma_p^2$
Resíduo	122	$Q_2 = 10516226,24$	$\sigma_t^2 + m\sigma_\delta^2 + mn\sigma_e^2$
Indivíduo (I) Parcela	558	$Q_3 = 7537364,96$	$\sigma_t^2 + m\sigma_\delta^2$
Medição (M)	11	-	-
(I x M)	8173	$Q_4 = 2563935,63$	σ_t^2
Total	8927		

Com base nos resultados da Tabela 6.17, obtêm-se as seguintes estimativas:

$$\hat{\sigma}_t^2 = Q_4 = 256393563$$

$$\hat{\sigma}_{\delta}^2 = (Q_e - Q_4)/m = 4144524441$$

$$\hat{\sigma}_e^2 = (Q_2 - Q_3)/(mn) = 6205961$$

$$\hat{\sigma}_p^2 = (Q_1 - Q_2)/(mnb) = 625326081$$

$$\hat{\sigma}_y^2 = 3102980292$$

$$\hat{\sigma}_a^2 = 4\hat{\sigma}_p^2 = 2501304323$$

$$\hat{h}_{ai}^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_y^2} = 0,0806 : \text{herdabilidade individual no sentido restrito.}$$

$$\hat{\rho}_i = \frac{\hat{\sigma}_p^2 + \hat{\sigma}_e^2 + \hat{\sigma}_{\delta}^2}{\hat{\sigma}_y^2} = 0,1737 : \text{repetibilidade individual.}$$

A repetibilidade para a seleção individual ao nível de médias de 12 safras mensais equivale a:

$$\hat{\rho}_{im}^2 = \frac{m \hat{\rho}_i}{1 + (m-1) \hat{\rho}_i} = \frac{12 \cdot 0,1737}{1 + (11 \cdot 0,1737)} = 0,7161$$

A herdabilidade para a seleção individual ao nível de médias (ou totais) de 12 safras mensais é dada por:

$$\hat{h}_{aim}^2 = \frac{m \hat{h}_{ai}^2}{1 + (m-1) \hat{\rho}_i} = \frac{12 \cdot 0,0806}{1 + (11 \cdot 0,1737)} = 0,3323$$

Por esta expressão, pode-se obter também que a herdabilidade para a seleção baseada na produção de 24 meses (2 anos) será de 0,387.

A eficiência do uso de safras mensais adicionais, em relação ao uso de apenas uma safra é dado por: $[m/[1+(m-1) \rho_i]]^{1/2}$. Assim, são as seguintes as eficiências (E), para o presente caso.

m	E
1	1,00
2	1,31
3	1,49
4	1,62
5	1,72
6	1,79
7	1,85
8	1,90
9	1,94
10	1,97
11	2,00
12	2,03
13	2,05
14	2,07
15	2,09

Verifica-se que o aumento na eficiência seletiva após 12 avaliações mensais é de apenas 2% para cada avaliação mensal a mais. Com 12 avaliações mensais, a acurácia seletiva equivale a $[\hat{h}_{aim}^2]^{1/2} = 0,5765$ e a acurácia máxima possível equivale a $\hat{h}_{ai} / \hat{\rho}_i = (0,0806/0,1737)^{1/2} = 0,6812$, avaliando-se infinitos meses. Dessa forma, 12 avaliações permitem atingir 85% da acurácia máxima possível. Para atingir 90% e 95% da acurácia máxima possível seriam necessárias 43 a 90 avaliações mensais, respectivamente, valores estes dados por $m = f(1 - \rho_i) / [(1 - f) \rho_i]$, em que f é a proporção da acurácia máxima possível (Resende et al., 1995a).

Repetibilidade da média de progênie ou clone

A repetibilidade da média de progênie, para uma safra mensal, é dada por:

$$\rho_f = \frac{COV(\bar{Y}_{i..m}, \bar{Y}_{i..m'})}{Var(\bar{Y}_{i..m})} = \frac{\sigma_p^2 + \sigma_e^2/b + \sigma_\delta^2/nb}{\sigma_p^2 + \sigma_e^2/b + \sigma_\delta^2/nb + \sigma_t^2/nb} = \frac{(n-1) t_1 + (b-1) n t_2 + \rho_i}{1 + (n-1) t_1 + (b-1) n t_2}$$

, em que:

$$t_2 = \rho_a h_a^2 + \rho_d (h_g^2 - h_a^2);$$

$$t_1 = c^2 + t_2.$$

Para m safras mensais, a repetibilidade da média de progênie é dada por:

$$\rho_{fm} = \frac{1 + (n-1) t_1 + (b-1) n t_2 - (1 - \rho_i)}{[1 + (m-1) \rho_i + (n-1) m t_1 + (b-1) n m t_2]/m} = \frac{m \rho_f}{1 + (m-1) \rho_f}.$$

Com $m = 12$, no presente exemplo da acerola, obtém-se:

$$\hat{\rho}_f = 0,355, \quad \hat{\rho}_{fm} = 0,869.$$

Alternativamente, este resultado pode ser obtido calculando a repetibilidade a partir de uma análise de variância de medidas repetidas, baseada em médias de progênies.

Neste caso, trabalhando-se ao nível de médias de progênies, a repetibilidade ao nível de indivíduos (ρ_i) pode ser obtida a partir da repetibilidade ρ_f por meio da expressão:

$$\rho_i = \rho_f + (\rho_f - 1)[(n-1) t_1 + (b-1) n t_2]$$

Assim, conhecendo-se ρ_f e a herdabilidade, obtém-se ρ_i .

Com base nos resultados apresentados, pode-se concluir:

- O emprego de modelos que ignoram a estrutura experimental superestimam a repetibilidade.
- Os modelos (1, 2 e 5) que utilizam a estrutura experimental e permitem a estimação simultânea da herdabilidade e da repetibilidade devem ser empregados no melhoramento de espécies perenes.

6.9. Estimativas de Repetibilidade em Espécies Perenes e Classificação de Magnitudes

Na Tabela 6.18 são apresentadas estimativas de repetibilidade individual em várias espécies vegetais perenes.

Tabela 6.18. Estimativas do parâmetro genético repetibilidade individual (ρ) em algumas espécies perenes.

Espécie	Caráter	$\hat{\rho}$	Nº. Estimativas	Referência
<i>Eucalyptus urophylla</i>	Circunferência do tronco	0,92	21	Resende (1999b)
	Incremento anual na circunferência	0,69	1	Resende (1999b)
Erva-mate	Massa foliar	0,70	3	Resende et al. (1995a)
Seringueira	Produção de borracha	0,80	4	Resende et al. (1996b)
Cacaueiro	Número de frutos sadios	0,50	1	Dias & Souza (1993)
	Peso de sementes úmidas por planta	0,53	1	Dias & Souza (1993)
	Peso de sementes úmidas por frutos	0,67	1	Dias & Souza (1993)
	Número de frutos por planta	0,57		Resende & Dias (1999)
Coqueiro	Número de frutos por planta	0,45	1	Siqueira (1982)
Cupuaçu	Comprimento do fruto	0,82	1	Costa et al. (1997)
	Circunferência do fruto	0,50	1	Costa et al. (1997)
	Espessura da casca	0,93	1	Costa et al. (1997)
	Peso do fruto	0,77	1	Costa et al. (1997)
	Número de sementes	0,55	1	Costa et al. (1997)
	Porcentagem de casca	0,63	1	Costa et al. (1997)
	Porcentagem de polpa	0,70	1	Fonseca et al. (1990))
	Porcentagem de sementes	0,72	1	Fonseca et al. (1990))
	Porcentagem de acidez da polpa	0,63	1	Fonseca et al. (1990)
	Porcentagem de casca	0,50	1	Fonseca et al. (1990)
	Peso da polpa	0,61	1	Fonseca et al. (1990)
	Relação brix/acidez	0,60	1	Fonseca et al. (1990)
	Peso da casca	0,58	1	Fonseca et al. (1990)
	Largura do fruto	0,56	1	Fonseca et al. (1990)
	Peso total	0,56	1	Fonseca et al. (1990)

Continua...

Tabela 6.18. Continuação.

Espécie	Caráter	$\hat{\rho}$	Nº. Estimativas	Referência
Guaraná	Tamanho da inflorescência	0,77	1	Valois et al. (1979)
	Número de botões	0,74	1	Valois et al. (1979)
	Número de frutos	0,66	1	Valois et al. (1979)
	Número de sementes	0,74	1	Valois et al. (1979)
Média		0,66		

De maneira geral, a classificação dos coeficientes de repetibilidade quanto à magnitude pode ser realizada comparando-se o ganho fenotípico permanente a ser obtido considerando uma medição (G_1) com aquele a ser obtido supondo-se m medições (G_m) de onde se obtém a relação $G_1/G_m = \{[1+(m-1) \rho] / m\}^{1/2}$. Considerando o padrão $m = 2$ para G_m , são as seguintes as classificações para a repetibilidade:

- repetibilidade alta: $G_1/G_m \geq 0,90 \Rightarrow \rho \geq 0,60$
- repetibilidade média: $0,80 < G_1/G_m < 0,90 \Rightarrow 0,30 < \rho < 0,60$
- repetibilidade baixa: $G_1/G_m \leq 0,80 \Rightarrow \rho \leq 0,30$

Com base nesta classificação, pode-se inferir que a maioria dos caracteres quantitativos de espécies vegetais perenes apresenta repetibilidade alta. O valor médio encontrado foi de 0,66, indicando que mesmo com \hat{h}_a^2 baixa (em torno de 0,20 para os referidos caracteres em erva-mate, seringueira, *Pinus* e *Eucalyptus*) a repetibilidade é alta. Considerando que h_g^2 situa-se ao redor de 0,30 para seringueira e para *Eucalyptus*, verifica-se que a fração $p^2 = \sigma_p^2 / \sigma_y^2$ provavelmente seja alta (em torno de 0,35 a 0,50) para espécies vegetais perenes, revelando um forte componente ambiental permanente. Em animais domésticos, as repetibilidades verificadas (Willis, 1991; Resende & Rosa-Perez ,1999b) não são tão altas, citando-se os valores de 0,45 para produção de leite ($h^2 = 0,30$) em bovinos e 0,15 para tamanho e peso da leitegada em suínos, de forma que o componente p^2 parece ser bem mais baixo. De fato, a superioridade de p^2 para espécies vegetais perenes em relação àquelas verificadas para os animais é esperada em função de os animais estarem menos sujeitos a efeitos ambientais permanentes devido a suas capacidades de movimento.

Para espécies como o coqueiro, que apresenta flutuações bianuais de produção (Shrikhande, 1957), as baixas estimativas de repetibilidade obtidas podem ser atribuídas ao tipo de análise empregada. Nestes casos, o método de análise de variância subestima a repetibilidade (Vasconcellos et al., 1985).

6.10. Tamanho Amostral e Precisão das Estimativas de Repetibilidade

Para a repetibilidade (ρ), o desvio padrão pode ser calculado considerando o estimador do desvio padrão do coeficiente de correlação linear simples, já que a repetibilidade é uma correlação intraclass. Assim:

$$\sigma_{\rho} = \frac{(1-\rho)[1+(m-1)\rho]}{[(1/2)m(m-1)(N-1)]^{1/2}} \text{ para o caso de mais de duas medições (m) (Turner \& Young, 1969) e: } \sigma_{\rho} = \frac{1-\rho^2}{(N-1)^{1/2}} \text{ para o caso de apenas duas medições.}$$

Considerando duas medições, o número (N) de indivíduos necessários para obter uma estimativa de ρ com um desejado desvio padrão σ_{ρ} é dado por: $N = (1-\rho^2)^2 / \sigma_{\rho}^2 + 1$. Na Tabela 6.19 são apresentados os valores adequados de N para alguns valores de ρ e σ_{ρ} .

Tabela 6.19. Valores adequados do número de indivíduos N a serem medidos, visando à obtenção de estimativas da repetibilidade (ρ) com determinado valor desejado de σ_{ρ} .

ρ	$\sigma_{\rho}=0,2\rho$	N	σ_{ρ}	N
0,10	0,02	2451	0,05	393
0,20	0,04	577	0,05	370
0,30	0,06	231	0,05	332
0,40	0,08	111	0,05	283
0,50	0,10	57	0,05	226
0,60	0,12	29	0,05	165
0,70	0,14	14	0,05	105
0,80	0,16	6	0,05	53
0,90	0,18	3	0,05	16

Considerando um desvio padrão de 0,05, verifica-se que de 100 a 400 indivíduos são suficientes para obter estimativas precisas de repetibilidade que apresentam valores paramétricos entre 0,70 e 0,10, respectivamente.

Capítulo 7

**Procedimentos Ótimos
de Seleção com Dados
Balanceados e Desbalanceados**

7.1. Introdução

Os procedimentos ótimos de predição de valores genéticos e seleção no melhoramento de espécies perenes podem ser resumidos em dois: índice usando todos os efeitos aleatórios do modelo estatístico (índice multiefeitos) para o caso balanceado; BLUP (melhor predição linear não viciada) individual para os casos balanceado e desbalanceado.

Neste capítulo é apresentada uma abordagem integrada sobre os dois procedimentos mencionados, para cada situação dos diferentes delineamentos experimentais e de cruzamento, presença de um e vários locais de experimentação, uma e várias populações (germoplasma), uma e várias medições por indivíduo. Inicialmente são apresentados aspectos fundamentais referentes à predição de variáveis aleatórias.

7.2. Predição de Variáveis Aleatórias e Metodologia de Modelos Mistos

O sucesso do melhoramento genético depende, sobretudo, da adoção de acurados procedimentos de seleção. A estruturação destes procedimentos baseia-se na estimativa de componentes de variância e predição de valores genéticos visando à avaliação genética dos candidatos à seleção.

A metodologia de modelos mistos foi derivada por Henderson (1949), no contexto da avaliação genética de gado de leite. Tal metodologia foi apresentada, formalmente, em 1973 (Henderson, 1973), mas passou a ser utilizada na prática a partir da década de 1980, com os avanços da tecnologia computacional. O procedimento de predição foi denominado BLUP, em 1973 (Henderson, 1973), para designar “melhor predição linear não viciada” de variáveis aleatórias. A expressão melhor predição linear não viciada já havia sido usada antes por Goldberger (1962), mas a sigla BLUP foi criada por Henderson.

O procedimento BLUP ajusta os dados para os efeitos ambientais identificáveis (tais quais os efeitos de bloco, do ano de medição, do local de plantio, dentre outros) e simultaneamente prediz os valores genéticos dos indivíduos.

Em termos cronológicos, os seguintes fatos podem ser relatados em relação à evolução do procedimento BLUP: (i) desenvolvimento conceitual do BLUP (Henderson, 1949); (ii) estimativa dos efeitos fixos por quadrados mínimos generalizados (Henderson et al., 1959); (iii) apresentação do BLUP em conexão com a teoria dos índices de seleção (Henderson, 1963); (iv) denominação e apresentação do BLUP em termos das equações de modelo misto (Henderson, 1973); (v) descoberta de fórmulas para escrever a inversa da matriz de parentesco diretamente (Henderson, 1976); (vi) criação do modelo animal (Henderson & Quaas, 1976); (vii) denominação do BLUP ao nível de indivíduo como BLUP – Modelo Animal (Quaas & Pollak, 1980); (viii) desenvolvimento do BLUP – Modelo Animal com efeito de grupo genético (Quaas, 1988; Westell et al., 1988).

Atualmente, com o advento da avaliação genética via inferência bayesiana (simulação estocástica por meio do amostrador de Gibbs), as equações de modelo misto de Henderson são ainda o principal veículo computacional. Em realidade, o BLUP pode também ser derivado sob o enfoque bayesiano (Gianola & Fernando, 1986), e, também, ser estendido para o contexto dos modelos lineares generalizados.

7.2.1. Predição de variáveis aleatórias

Considere a e y , vetores de variáveis aleatórias, conjuntamente distribuídas, em que y é observável (dados) e a não observável (valores genéticos). Para predição de a a partir de y , três procedimentos de predição são disponíveis: melhor predição (BP), melhor predição linear (BLP) e melhor predição linear não viciada (BLUP). O procedimento BP é disponível quando são conhecidos todos os parâmetros da distribuição conjunta de y e a , isto é, quando se conhece $f(y,a)$. Os procedimentos BLP e BLUP são melhores (best) em situações em que se conhecem alguns dos parâmetros de $f(y,a)$, mas não todos. Para o BLP, os primeiros (médias) e segundos (variâncias) momentos dos dados são assumidos como conhecidos, enquanto que para o BLUP, somente os segundos momentos são assumidos como conhecidos.

Assim, embora de maneira genérica, predição refira-se a eventos futuros, significa, também, a estimativa da realização de uma variável aleatória amostrada de uma população com uma conhecida estrutura de variância-covariância.

7.2.2. BP

Para predição de a a partir de y , deve-se encontrar uma função ($f(y)$) de y tal que $E(\hat{a}-a)$ seja mínimo. Esta função é $E(a|y)$, ou seja, o melhor preditor de a é a média condicional de a dado y . Entretanto, para a derivação desta média condicional, torna-se necessário o conhecimento da distribuição conjunta de a e y . Assim, considerando distribuição normal multivariada entre a e y é o seguinte o preditor:

$$\hat{a} = E(a|y) = \mu_a + CV^{-1}(y - \mu_y), \text{ em que:}$$

$$\mu_a = E(a) : \text{média de } a$$

$$\mu_y = E(y) : \text{média de } y$$

C e V : matrizes de covariância entre a e y e de variância de y , respectivamente.

Os valores μ_a , μ_y , C e V , devem ser conhecidos ou, na prática, estimados de maneira fidedigna.

As propriedades dos preditores BP são (Henderson, 1984):

(i) não vício: $E(\hat{a})=E(a)$

$$(ii) \text{Var}(\hat{a}_i - a_i) = \text{Var}(a|y)$$

(iii) correlação $r_{\hat{a}a}$ máxima para todas as funções de y .

A forma do preditor BP pode ser linear ou não linear, ou seja, não necessariamente linear em y (Schaeffer & Mao, 1988; Searle et. al. 1992).

O preditor BLP (apresentado no tópico seguinte) é exatamente o mesmo que o preditor BP, porém, é derivado sem a suposição de normalidade. Assim, o preditor BP somente difere do BLP pela suposição adicional de normalidade multivariada (Ronningen & Van Vleck, 1985; Searle et. at. 1992).

Segundo Henderson (1977, 1984), um preditor BP só poderia ser aplicado eficientemente na prática, em situações em que os candidatos à seleção apresentassem iguais quantidades de informações e fossem não correlacionados. Nesta situação, os mesmos ponderadores das informações fenotípicas seriam utilizados para todos os candidatos à seleção.

Assim, em resumo, os três procedimentos e situações em que devem ser aplicados podem assim ser classificados:

BP : iguais quantidades e precisões das informações associadas a todos os candidatos à seleção, primeiros momentos (médias) e segundos momentos (variâncias) conhecidos.

BLP : diferentes quantidades e precisões das informações associadas aos candidatos à seleção, primeiros e segundos momentos conhecidos.

BLUP : diferentes quantidades e precisões das informações associadas aos candidatos à seleção, segundo momentos conhecidos, primeiros momentos não conhecidos.

7.2.3. BLP

Considerando agora preditores que são lineares em y , da forma $\hat{a} = a + By$, para um vetor a e matriz B , tem-se (sem qualquer suposição de normalidade):

$$\hat{a} = \mu_a + CV^{-1}(y - \mu_y).$$

Assim, restringindo o preditor a ser linear, a forma da distribuição de y não necessita ser conhecida, mas apenas os primeiros e segundo momentos. Quando y tem uma distribuição normal multivariada, então BLP é igual ao BP. O BLP tem o menor erro quadrático médio dentre todos os preditores lineares de a .

7.2.4. BLUP

Considerando μ_y não conhecido, o preditor é forçado a ser não viciado e o melhor preditor linear não viciado é:

$\hat{a} = \mu_a + CV^{-1}(y - X\hat{\beta})$, em que:

$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$: estimador de quadrados mínimos generalizados (GLS) ou melhor estimador linear não viciado (BLUE) do vetor (β) de efeitos fixos.

X : matriz de incidência para β .

Assim, o preditor BLUP é o mesmo do BLP, porém substituindo μ_y por $X\hat{\beta}$. O BLUP tem o menor erro quadrático médio dentre todos os preditores lineares não viciados de a .

No BLUP, não víncio é um critério do procedimento de predição e não uma consequência como no BP e BLP. Introduzindo-o como critério, surge $\hat{\beta}$ no preditor.

As diferenças entre BP, BLP e BLUP são sutis, porém importantes estatisticamente (Schaeffer & Mao, 1988). É importante relatar que, na prática, os primeiros e segundos momentos não são conhecidos, mas sim estimados dos dados amostrais. Assim, a rigor, nas aplicações práticas não se tem BLP e BLUP, mas sim EBLP e EBLUP, significando BLP e BLUP empíricos, respectivamente, devido à substituição dos valores paramétricos (primeiros e segundos momentos no caso do BLP e segundos momentos no caso do BLUP) por suas estimativas.

Conforme Dempfle (1989), no preditor BLP não é especificado o que fazer com os primeiros momentos ($E(y)$). A literatura tem tratado como BLP os preditores utilizados quando se estimam os primeiros momentos por quadrados mínimos ordinários (Ronningen e Van Vleck, 1985; White e Hodge, 1989), embora tal tratamento não seja estatisticamente exato.

7.2.5. Modelos lineares e predição de valores genéticos

O vetor de dados fenotípicos pode ser expresso pelo seguinte modelo estatístico:

$y = X\beta + Za + \varepsilon$, em que:

μ : vetor de médias associadas aos efeitos aleatórios

β : vetor de efeitos fixos

a : vetor aleatório de valores genéticos

ε : vetor aleatório de erros (resíduos)

X, Z : matrizes de incidência para β e a , respectivamente.

Ao vetor β estão associados os efeitos ambientais identificáveis, como por exemplo o efeito de locais, ano de avaliação, e, também, o efeito da média geral da população. Estes efeitos são quantificáveis e utilizados para o ajuste dos dados por meio de $(y - X\beta)$. O vetor ε contém efeitos ambientais não identificáveis e, portanto, não quantificáveis, pois ocorrem aleatoriamente. Assim, na

predição de a deve-se adotar um procedimento que permita isolar ε da fração $(y - X\beta) = (Za + \varepsilon)$.

Expressos como desvios da média geral, os valores genéticos podem ser preditos genericamente (Henderson, 1984):

$$\hat{a} = CV^{-1}(y - E(y)) = GZV^{-1}(y - E(y)), \text{ em que :}$$

C : matriz de covariância entre a e y .

G : matriz de covariância de a .

V : matriz de covariância de y .

$E(y)$: vetor de valores esperados de y , equivalendo a $E(y) = X\beta$.

A seguir são apresentados alguns resultados em modelos lineares e suas provas algébricas.

7.2.5.1. Modelo linear de efeitos fixos

(a) Estimadores de quadrados mínimos (LS)

Modelo linear

$$y = X\beta + e, \quad v(e) = I\sigma^2$$

Estimador

$$(X'X) \hat{\beta} = X'y$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

Definições

y : vetor de observações

β : vetor de efeitos desconhecidos

e : vetor de erros

X : matriz de incidência conhecida

Prova

Os estimadores de quadrados mínimos de β são aqueles que minimizam a soma (S) de quadrados dos erros de estimação. Assim, tem-se:

$$\begin{aligned}
 S &= \sum (y - \hat{y})^2 = (y - \hat{y})'(y - \hat{y}) \\
 &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
 &= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'\hat{\beta} \\
 \frac{\partial S}{\partial \hat{\beta}} &= -2X'y + 2X'X\hat{\beta} = 0 \\
 (X'X)\hat{\beta} &= X'y \\
 \hat{\beta} &= (X'X)^{-1}X'y
 \end{aligned}$$

(b) Estimadores de quadrados mínimos generalizados (GLS)

Modelo linear

$$y = X\beta + e, \quad v(e) = V$$

Estimador

$$\begin{aligned}
 (X'V^{-1}X)\hat{\beta} &= X'V^{-1}y \\
 \hat{\beta} &= (X'V^{-1}X)^{-1}X'V^{-1}y
 \end{aligned}$$

Definição

V : matriz de covariância dos erros.

Prova

Os estimadores de quadrados mínimos generalizados de β são aqueles que minimizam a soma (S) ponderada dos quadrados dos erros de estimação. Assim:

$$\begin{aligned}
 S &= (y - \hat{y})'V^{-1}(y - \hat{y}) \\
 &= (y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}) \\
 &= y'V^{-1}y - 2\hat{\beta}'X'V^{-1}y + \hat{\beta}'X'V^{-1}X\hat{\beta} \\
 \frac{\partial S}{\partial \hat{\beta}} &= -2X'V^{-1}y + 2X'V^{-1}X\hat{\beta} = 0 \\
 (X'V^{-1}X)\hat{\beta} &= X'V^{-1}y
 \end{aligned}$$

Nos casos em que V é diagonal, ou seja, quando os elementos da diagonal V são diferentes (heterocedasticidade) e os elementos fora da diagonal de V são zero (erros não correlacionados), o estimador $\hat{\beta}$ é denominado de estimador de quadrados mínimos ponderados.

c) Propriedades dos estimadores GLS

Não vício

$$E(\hat{\beta}|\beta) = \beta$$

Prova

$$\begin{aligned} E(\hat{\beta}) &= E((X'V^{-1}X)^{-1}X'V^{-1}y) \\ &= E[(X'V^{-1}X)^{-1}X'V^{-1}(X\beta + e)] \\ &= E((X'V^{-1}X)^{-1}X'V^{-1}X\beta) + E((X'V^{-1}X)^{-1}X'V^{-1}e) \\ &= E(I\beta) + 0 \\ &= \beta \end{aligned}$$

Variância do erro de estimação

$$Var(\hat{\beta} - \beta) = (X'V^{-1}X)^{-1}$$

Prova

$$\begin{aligned} Var(\hat{\beta} - \beta) &= Var(\hat{\beta}) \text{ uma vez que } \beta \text{ é fixo.} \\ &= V((X'V^{-1}X)^{-1}X'V^{-1}y) \\ &= (X'V^{-1}X)^{-1}X'V^{-1}V(y)V^{-1}X(X'V^{-1}X)^{-1} \\ &= (X'V^{-1}X)^{-1}X'V^{-1}V V^{-1}X(X'V^{-1}X)^{-1} \\ &= (X'V^{-1}X)^{-1} \end{aligned}$$

7.2.5.2. Modelo linear de efeitos aleatórios

(a) Melhor predição linear (BLP)

(a.1.) BLP de uma variável única

O BLP de uma variável aleatória única (não de um vetor) (g_i) é:

$$\hat{g}_i = b'y$$

em que:

$$Vb = c$$

$$V = Var(y)$$

$$c = cov(y, g_i)$$

Prova

O preditor BLP é aquele que minimiza a variância do erro de predição, denotada por $Q = \text{Var}(\hat{g}_i - g_i)$. Assim:

$$Q = \text{Var}(\hat{g}_i - g_i) = E(\hat{g}_i - g_i)^2 \text{ porque } E(\hat{g}_i) = E(g_i) = 0$$

$$= E(\hat{g}_i^2 - 2g_i\hat{g}_i + g_i^2)$$

$$= E((b'y)^2 - 2g_i b'y + g_i^2)$$

$$= b'Vb - 2b'c + \text{Var}(g_i)$$

$$\frac{\partial Q}{\partial b} = 2Vb - 2c = 0$$

$$Vb = c$$

(a.2.) BLP de um vetor aleatório

O BLP de um vetor de variáveis aleatórias (g) é:

$$\hat{g} = B'y$$

em que:

$$VB = C$$

$$V = \text{Var}(y)$$

$$C = \text{cov}(g, y)$$

Prova

O BLP de um único elemento de g (g_i) é:

$$\hat{g}_i = b_i'y$$

$$\text{Assim: } \hat{g} = B'y$$

em que:

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}; \quad C = \begin{bmatrix} \text{cov}(y, g_1) \\ \text{cov}(y, g_2) \\ \vdots \\ \text{cov}(y, g_n) \end{bmatrix} = \text{cov}(y, g)$$

O BLP de g minimiza Q da seguinte forma:

$$\begin{aligned} Q &= E \left[\sum_i (\hat{g}_i - g_i)^2 \right] = E(\hat{g} - g)'(\hat{g} - g)' \\ &= E(\hat{g}'\hat{g} - 2g'\hat{g} + g'g) \\ &= E(y'BB'y - 2g'B'y + g'g) \\ &= tr(BBV) - 2tr(B'C) + E(g'g) \\ \frac{\partial Q}{\partial B} &= 2VB - 2C = 0 \\ VB &= C \end{aligned}$$

(b) Modelos e expressões alternativas para o BLP

(b.1.) Modelo aleatório

$$y = g + e$$

Definições

y : vetor de observação.

g : vetor de variáveis aleatórias (valores genéticos).

e : vetor de erros.

$$E(y) = E(g) = E(e) = 0$$

$$Var \begin{bmatrix} y \\ g \\ e \end{bmatrix} = \begin{bmatrix} V & G & R \\ G & G & O \\ R & O & R \end{bmatrix} \quad V = G + R$$

Predictor BLP de g :

$$(R^{-1} + G^{-1}) \hat{g} = R^{-1}y$$

Prova

Da expressão convencional do BLP e, considerando $V = G + R$ e $\text{COV}(y, g) = G$, tem-se:

$$\hat{g} = B'y$$

em que:

$$VB = C$$

$$(G + R) B = G$$

$$B = (G + R)^{-1} G$$

$$\hat{g} = G(G + R)^{-1} y$$

$$= (I + G^{-1} R)^{-1} y$$

$$= (R^{-1} + G^{-1})^{-1} R^{-1} y$$

$$(R^{-1} + G^{-1}) \hat{g} = R^{-1} y$$

(b.2.) Modelo aleatório

$$y = Zg + e$$

Definições

Z : matriz de incidência conhecida.

$$Var \begin{bmatrix} y \\ g \\ e \end{bmatrix} = \begin{bmatrix} V & ZG & R \\ GZ & G & O \\ R & O & R \end{bmatrix} \quad V = ZGZ + R$$

Preditor BLP de g

$$(Z' R^{-1} Z + G^{-1}) \hat{g} = Z' R^{-1} y$$

Prova

Demonstração de uma equivalência (1)

$$V^{-1} = (ZGZ + R)^{-1} = R^{-1} - R^{-1} Z (Z' R^{-1} Z + G^{-1})^{-1} Z' R^{-1}, \text{ pois:}$$

$$\begin{aligned} & (ZGZ + R)(R^{-1} - R^{-1} Z (Z' R^{-1} Z + G^{-1})^{-1} Z' R^{-1}) = \\ & = (ZGZ + R)(I - R^{-1} Z (Z' R^{-1} Z + G^{-1})^{-1} Z') R^{-1} \\ & = (ZGZ + R - ZGZ R^{-1} Z (Z' R^{-1} Z + G^{-1})^{-1} Z' - Z (Z' R^{-1} Z + G^{-1})^{-1} Z') R^{-1} \\ & = (ZGZ + R - ZG(Z' R^{-1} Z + G^{-1})(Z' R^{-1} Z + G^{-1})^{-1} Z') R^{-1} \\ & = (ZGZ + R - ZGZ) R^{-1} = I \end{aligned}$$

Usando $V = ZGZ' + R$ e $C = ZG$ no BLP convencional e a demonstração (1), tem-se que:

$$\hat{g} = B'y$$

$$B = V^{-1}C = (ZGZ + R)^{-1}ZG$$

$$\hat{g} = GZ'(ZGZ + R)^{-1}y$$

$$= GZ'(R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1})y$$

$$= G(I - Z'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1})Z'R^{-1}y$$

$$= G[I + G^{-1}(Z'R^{-1}Z + G^{-1})^{-1} - G^{-1}(Z'R^{-1}Z + G^{-1})^{-1} - Z'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}]Z'R^{-1}y$$

$$= G[I + G^{-1}(Z'R^{-1}Z + G^{-1})^{-1} - (Z'R^{-1}Z + G^{-1})(Z'R^{-1}Z + G^{-1})^{-1}]Z'R^{-1}y$$

$$= (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}y$$

c) Propriedades dos estimadores BLP

$Var(\hat{g}) = COV(\hat{g}, g) = CV^{-1}C = G - (Z'R^{-1}Z + G^{-1})^{-1}$ de forma que a regressão de g em $\hat{g} = 1$.

Prova

$$\hat{g} = B'y$$

$$Var(\hat{g}) = B'VB \text{ (pois } B = V^{-1}C\text{)}$$

$$= C'V^{-1}C$$

$$Cov(\hat{g}, g) = Cov(B'y, g)B$$

$$= B'C$$

$$= C'V^{-1}C$$

$$C'V^{-1}C = GZ(ZGZ + R)^{-1}ZG \\ = G - (Z'R^{-1}Z + G^{-1})^{-1} \text{ (usando a demonstração 1).}$$

Os preditores BLP apresentam as seguintes propriedades:

- Maximização da correlação entre os valores genéticos preditos e os verdadeiros, ou seja, maximização da acurácia.
- Minimização do quadrado da diferença entre os valores genéticos preditos e os verdadeiros, ou seja, minimização da variância do erro de previsão (requisito para um procedimento “Best”: $E(\hat{a} - a)^2 \rightarrow \min$).
- Maximização da probabilidade de selecionar o melhor entre dois indivíduos quaisquer.
- Maximização da probabilidade de selecionar o melhor entre vários indivíduos.

- Maximização do ganho genético por ciclo de seleção.

Estas três últimas propriedades são válidas na suposição de distribuição conjunta normal multivariada de y e a .

A variância do erro de predição pode ser calculada por $PEV: G - C'V^{-1}C$ e a acurácia por $r_{aa} = [(C'V^{-1}C/G)]^{1/2}$.

7.2.5.3. Modelos lineares mistos com efeitos fixos e aleatórios

(a) Equações de modelo misto de Henderson

Modelo misto

$$y = X\beta + Zg + e$$

Equações de modelo misto

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

A estimativa GLS de β é a solução $\hat{\beta}$ das equações de modelo misto.

Prova

O estimador GLS de β é: $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$; $V = ZGZ + R$

Das equações de modelo misto, tem-se:

$$\hat{g} = (Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}y - Z'R^{-1}X\hat{\beta})$$

Assim:

$$X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{g} = X'R^{-1}y$$

$$X'R^{-1}X\hat{\beta} + X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}y - Z'R^{-1}X\hat{\beta}) = X'R^{-1}y$$

$$(X'R^{-1}X - X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X)\hat{\beta} = X'R^{-1}y - X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}y$$

$$X'DX\hat{\beta} = X'Dy$$

em que:

$$D = R^{-1} - R^{-1}Z(Z'R^{-1}Z + A^{-1})^{-1}Z'R^{-1}$$

Pela demonstração (1) $D = V^{-1}$, de forma que $X'DX\hat{\beta} = X'Dy$ é equivalente ao GLS de β , dado por $X'V^{-1}X\hat{\beta} = X'V^{-1}y$.

(b) GLS + BLP

Modelo misto

$$y = X\beta + Zg + e$$

O vetor g pode ser obtido pela correção de y usando a estimativa GLS de β e posteriormente aplicando o BLP, da seguinte maneira:

$$\hat{g} = B'(y - X\hat{\beta})$$

$$\hat{g} = CV^{-1}(y - X\hat{\beta}), \text{ em que: } V = ZGZ + R ; \quad C = ZG$$

$$\hat{g} = GZ(ZGZ + R)^{-1}(y - X\hat{\beta})$$

$$\hat{g} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\hat{\beta}), \text{ conforme demonstrado no item (b.2.).}$$

O vetor g predito por GLS + BLP é idêntico ao obtido pela solução das equações de modelo misto.

Prova

Das equações de modelo misto, obtém-se:

$$\begin{aligned}\hat{g} &= (Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}y - Z'R^{-1}X\hat{\beta}) \\ &= (Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1})(y - X\hat{\beta})\end{aligned}$$

Como $\hat{\beta}$ obtido pelas equações de modelo misto e por GLS são idênticos (item anterior), demonstra-se a equivalência anunciada.

(c) Melhor predição linear não viciada

Modelo misto

$$y = X\beta + Zg + e$$

O preditor linear de g ($\hat{g} = B'y$) o qual é não viciado [$E(\hat{g}) = 0$] e com mínima variância do erro de predição é a solução do sistema:

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

Prova

$$\begin{aligned}E(\hat{g}) &= E(B'y) \\ &= B'X\beta = 0 \text{ para todo } \beta\end{aligned}$$

Isto é, $B'X = 0$

$$\begin{aligned}
tr[Var(\hat{g} - g)] &= E(\hat{g} - g)'(\hat{g} - g) \\
&= E(B'y - g)'(B'y - g) \\
&= E(y'BB'y - 2g'B'y + g'g) \\
&= tr(BB'(ZGZ+R)) - 2tr(B'ZG) + tr(G)
\end{aligned}$$

Para minimizar $tr[Var(\hat{g} - g)]$ sujeito a $B'X = 0$, o multiplicador de Lagrange pode ser usado:

$$\begin{aligned}
&\frac{\partial}{\partial B} tr[BB'(ZGZ+R)] - 2tr(B'ZG) + tr(G) + \sum_i (b_i'X - 0) \lambda_i \\
&= 2(ZGZ+R)B - 2ZG + X\Lambda = 0, \text{ em que: } \Lambda = \{\lambda_i\} \\
&B = (ZGZ+R)^{-1}(ZG - \frac{1}{2}X\Lambda)
\end{aligned}$$

Usando $X'B = 0$

$$\begin{aligned}
&X'(ZGZ+R)^{-1}(ZG - \frac{1}{2}X\Lambda) = 0 \\
&X'(ZGZ+R)^{-1}ZG = \frac{1}{2}X'(ZGZ+R)^{-1}X\Lambda \\
&\frac{1}{2}\Lambda = (X'V^{-1}X)^{-1}X'V^{-1}ZG, \text{ em que: } V = ZGZ+R
\end{aligned}$$

Assim:

$$\begin{aligned}
B &= (ZGZ+R)^{-1}(ZG - X(X'V^{-1}X)^{-1}X'V^{-1}ZG) \\
\hat{g} &= B'y \\
&= (G'Z' - GZV^{-1}X(X'V^{-1}X)^{-1}X')V^{-1}y \\
&= GZ'(I - V^{-1}X(X'V^{-1}X)^{-1}X')V^{-1}y \\
&= GZ'(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X)V^{-1}y \\
&= GZV^{-1}(y - X\hat{\beta}) \\
&= (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\hat{\beta}) \quad (\text{conforme item b}).
\end{aligned}$$

Assim, BLUP de g equivale a GLS + BLP de g .

(d) Máximo a posteriori de g

Sendo $P(g|y)$ = probabilidade de g dado y , o máximo a posteriori (MAP) de g é dado pela maximização de $P(g|y)$.

Quando g e y seguem uma distribuição normal multivariada, o MAP de g é equivalente ao BLUP de g .

Prova

Se $Z \sim N(\mu, V)$, ou seja, $P(Z) = \frac{1}{(2\pi)^{n/2}|V|^{n/2}} e^{-\frac{1}{2}(y-\mu)V^{-1}(y-\mu)}$ em que $n = \text{ordem de } y$, tem-se usando o teorema de Bayes:

$$\begin{aligned} P(g|y) &= \frac{P(y|g) P(g)}{P(y)} \\ \log P(g|y) &= \log P(y|g) + \log P(g) - P(y) \\ &= -\frac{n}{2} \log|R| - \frac{1}{2}(y - X\beta - Zg)'R^{-1}(y - X\beta - Zg) - \frac{1}{2} \log|G| - \frac{n}{2} g'A^{-1}g + \text{constante} \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \beta} X'R^{-1}(y - X\hat{\beta} - Z\hat{g}) &= 0 & \Rightarrow X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{g} &= X'R^{-1}y \\ \frac{\partial}{\partial g} Z'R^{-1}(y - X\hat{\beta} - Z\hat{g}) - G^{-1}g &= 0 & \Rightarrow Z'R^{-1}X\hat{\beta} + (Z'R^{-1}Z + G^{-1})\hat{g} &= Z'R^{-1}y \end{aligned}$$

Esta última expressão é equivalente ao BLUP de g .

(e) Variâncias dos erros de predição

Definindo:

$$C = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}, \text{ a qual equivale à matriz dos coeficientes das equações de modelo misto, a variância do erro de predição de } \hat{\beta} \text{ e de } \hat{g} \text{ é dada por:}$$

$$Var \begin{pmatrix} \hat{\beta} & - & \beta \\ \hat{g} & - & g \end{pmatrix} = C^{-1}$$

Prova

$$\text{Fazendo-se } W = [XZ], u = \begin{bmatrix} \beta \\ g \end{bmatrix}, \text{ tem-se: } y = Wu + e = X\beta + Zg + e$$

As equações de modelo misto são:

$$(WR^{-1}W + \Lambda) \hat{\mu} = WR^{-1}y, \text{ em que } \Lambda = \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix}; C = WR^{-1}W + \Lambda;$$

$$Var(u) = \begin{pmatrix} 0 & 0 \\ 0 & G \end{pmatrix} = \Lambda^-, \text{ que é a inversa generalizada de } \Lambda.$$

Assim:

$$\begin{aligned}
 Var(\hat{u} - u) &= V[C^{-1}W'R^{-1}y - u] \\
 &= V[C^{-1}W'R^{-1}(Wu + e) - u] \\
 &= V[C^{-1}W'R^{-1}W - I]u + C^{-1}W'R^{-1}e \\
 &= V[C^{-1}(W'R^{-1}W - C)u + C^{-1}W'R^{-1}e] \\
 &= V[C^{-1}\Lambda u + C^{-1}W'R^{-1}e] \\
 &= C^{-1}\Lambda\Lambda^-\Lambda C^{-1} + C^{-1}W'R^{-1}RR^{-1}WC^{-1} \\
 &= C^{-1}\Lambda C^{-1} + C^{-1}W'R^{-1}WC^{-1} \\
 &= C^{-1}(\Lambda + W'R^{-1}W)C^{-1} \\
 &= C^{-1}
 \end{aligned}$$

A porção de C^{-1} , correspondente aos efeitos fixos equivale a:

$$\begin{aligned}
 Var(\hat{\beta} - \beta) &= [X'(R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1})X]^{-1} \\
 &= [X'V^{-1}X]^{-1}, \text{ usando o resultado (1).}
 \end{aligned}$$

Esta é a fórmula usual para a variância das estimativas de quadrados mínimos generalizados. $Var(\hat{\beta} - \beta) = Var(\hat{\beta})$, pois β é um efeito fixo com $Var(\beta) = 0$.

7.2.6. Preditores BLP vs BLUP

O método BLP, se associado a grandes desbalanceamentos, torna-se viável, principalmente em situações em que os valores genéticos são correlacionados. O método GLS, para a estimação BLUE de β , leva em consideração a influência do parentesco (correlação) entre os candidatos à seleção na estimação dos efeitos fixos, por meio da matriz V .

As principais diferenças entre BLUP e BLP são (Dempfle, 1989):

- (i) o BLP não indica o que fazer com $E(y)$.
 - (ii) o BLP não indica o que fazer na seleção envolvendo várias populações.
- Neste caso, a seleção pelo BLUP é baseada em $\hat{a}^* = P_1 \hat{\beta} + P_2 \hat{a}$, ou seja, em uma função estimável de $\hat{\beta}$ e \hat{a} .

Os preditores BLUP possuem todas as propriedades desejáveis dos preditores BLP e mais:

- a predição dos efeitos aleatórios é não viciada, assim como a estimação de qualquer função estimável dos efeitos fixos [$E(\hat{a}) = E(a) = 0$].

- considerando a distribuição normal multivariada, $E(a|\hat{a})$ é linear e, então, $E(a|\hat{a}) = E(\hat{a} + \varepsilon) = \hat{a}$. Esta propriedade é muito útil, pois significa que a média dos verdadeiros valores genéticos dos indivíduos selecionados equivale à média dos seus valores genéticos preditos.

Para a seleção envolvendo caracteres isolados, a propriedade de correto ordenamento dos candidatos à seleção é geralmente robusta a erros nas estimativas dos componentes de variância usados em lugar de seus valores paramétricos (isto não é, necessariamente, verdadeiro para a seleção multivariada envolvendo mais de um caráter ao mesmo tempo). A teoria genética enfatiza o uso de uma proporção entre componentes de variância (no caso, a herdabilidade individual ou uma função dela), a qual tem maior propriedade de constância do que cada componente de variância isolado (Henderson, 1979). Preditores não viciados podem ser obtidos mesmo quando componentes de variância incorretos são utilizados (Kackar & Harville, 1981). Descrições relevantes do BLUP são também apresentadas por Nicholas (1987) e Robinson (1991).

7.2.7. Álgebra linear numérica e aspectos computacionais da metodologia de modelos mistos

A obtenção de $\hat{\beta}$ e \hat{a} implica inversão da matriz C, dos coeficientes das equações de modelo misto. Isto é proibitivo dadas as dimensões desta matriz, que, na prática apresentam ordem muito elevada. Nesta situação, recomenda-se a resolução do sistema de equações lineares por meio de métodos iterativos a partir de vetores iniciais β_0 e a_0 fornecidos pelo usuário, não havendo, portanto, a necessidade de inversão da matriz C.

Considerando $R = I\sigma_e^2$ e definindo $\lambda = \sigma_e^2 / \sigma_a^2$, as equações de modelo misto passam a ser dadas por

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}.$$

Uma vez que a matriz A^{-1} pode ser escrita diretamente, ou seja, sem a necessidade de sua construção e posterior inversão, o sistema de equações correspondente às equações de modelo misto pode ser resolvido iterativamente, sem a necessidade de qualquer inversão de matrizes.

Os principais métodos iterativos de resolução de sistemas de equações lineares são relatados a seguir, sucintamente:

- Jacobi e Gauss-Seidel: adequados quando os elementos da diagonal da matriz dos coeficientes são dominantes. O método de Jacobi converge mais lentamente que o de Gauss-Seidel.
- Iteração em Bloco: adequado quando os elementos da diagonal da matriz dos coeficientes não são dominantes. Após a partição da matriz dos coefi-

entes em blocos, qualquer dos métodos iterativos relatados pode ser utilizado.

- Gradientes conjugados: superior aos demais para matrizes mais densas (menos esparsas).

Os métodos de Jacobi e de Gauss-Seidel são muito utilizados para modelos univariados. Para os modelos multivariados, o método de iteração em bloco é mais eficiente, pois os elementos da diagonal da matriz dos coeficientes podem não ser dominantes. Recentemente, o método dos gradientes conjugados vem sendo utilizado com vantagem (Strandén & Lidauer, 1999) para a solução de sistemas de equações lineares de grande dimensão.

Dada a importância do método numérico de Gauss-Seidel para a resolução iterativa de sistemas de equações lineares, este será descrito empregando um pequeno exemplo.

Seja o sistema de equações lineares:

$$\begin{cases} 4X_1 + X_2 + X_3 = 5 \\ -2X_1 + 5X_2 + X_3 = 0 \\ 3X_1 + X_2 + 6X_3 = -6,5 \end{cases}$$

As soluções para as três incógnitas X_1 , X_2 e X_3 são dadas por:

$$X_1^k = \frac{(5 - X_2^{k-1} - X_3^{k-1})}{4}; \quad X_2^k = \frac{(0 + 2X_1^k - X_3^{k-1})}{5}; \quad X_3^k = \frac{(-6,5 - 3X_1^k - X_2^k)}{6},$$

em que k , refere-se à k -ésima iteração.

Partindo-se de um vetor inicial $X^0 = (0, 0, 0)$, tem-se a 1ª iteração:
 $X_1^1 = \frac{(5 - 0 - 0)}{4} = \frac{5}{4}; \quad X_2^1 = \frac{(0 + 2 \cdot 0 - 0)}{5} = \frac{0}{5} = 0; \quad X_3^1 = \frac{(-6,5 - 3 \cdot 0 - 0)}{6} = -1,7967.$

Na 2ª iteração, tem-se:

$$X_1^2 = \frac{(5 - 1/2 + 1,7967)}{4} = 1,58; \quad X_2^2 = \frac{(0 + 2 \cdot 1,58 - (-1,7967))}{5} = 0,992; \quad X_3^2 = \frac{(-6,5 - 3 \cdot 1,58 - 0,992)}{6} = 2,03$$

O procedimento prossegue até que o menor valor de $|X^k - X^{k-1}| \leq \varepsilon$, em que ε é o erro desejado (geralmente $\leq 10^{-5}$).

7.2.8. Derivações alternativas do BLUP

A predição dos valores genéticos (a), a partir dos dados fenotípicos (y), baseia-se na regressão de a em y , dada por:

$$E(a|y) = \int a f(y, a) da / \int f(y, a) da, \text{ em que:}$$

$f(y, a)$: função densidade da distribuição de probabilidade conjunta de y e a .

Bulmer (1980) utilizou um procedimento em dois estágios para obtenção da regressão de a em y , dado por:

- (i) correção dos dados para os efeitos fixos ou obtenção de $y - X\hat{\beta}$, em que $\hat{\beta}$ é obtido por quadrados mínimos generalizados.
- (ii) obtenção de $E(a|y)$ por:

$$E(a|y) = \frac{COV[a, (y - X\hat{\beta})]}{Var(y - X\hat{\beta})} (y - X\hat{\beta}) = \hat{a}_B$$

Dois anos após, Gianola & Goffinet (1982) demonstraram a equivalência entre o preditor de Bulmer (\hat{a}_B) e o preditor de Henderson ($\hat{a}_H = CV^{-1}(y - X\hat{\beta})$).

Outra derivação da metodologia BLUP, sob o enfoque bayesiano, pode ser atribuída a Robertson (1955) e baseia-se na combinação de dois estimadores (fontes de informação) independentes. Neste caso, as equações de modelo misto (denominadas equações de modelo misto de Robertson) são dadas por:

$$\begin{bmatrix} X'R^{-1}X + S^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} E(\beta|y) \\ E(a|y) \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y + S^{-1} r_1 \\ Z'R^{-1}Y + G^{-1} 0 \end{bmatrix}, \text{ em que: } r_1 = E(\beta) \text{ e } 0 = E(a).$$

Considerando $S^{-1} \rightarrow 0$, este sistema de equações equivale às equações de modelo misto de Henderson. Detalhes a respeito da derivação das equações de modelo misto de Robertson são apresentados no Capítulo 12.

É importante mencionar, também, que o preditor de Henderson equivale exatamente ao preditor de Lush (Lush, 1931) se for ignorada a forma de estimação de β . Hazel & Terrill (1945) foram os primeiros a atentarem para a estimação de β a partir de dados desbalanceados.

7.2.9. Matriz de parentesco genético aditivo e de dominância

7.2.10. Matriz de parentesco genético aditivo e inversa

A matriz de parentesco A é essencial na avaliação genética pelo BLUP, permitindo a obtenção de resultados mais acurados (pela consideração das informações de parentes) e não viciados (pela consideração da correlação entre os valores genéticos, na estimação dos efeitos fixos).

A matriz A é simétrica e seus elementos da diagonal (a_{ii}) para cada indivíduo i é igual a $(1+F_i)$, em que F_i é o coeficiente de endogamia do indivíduo i . Os elementos fora (a_{ij}) da diagonal referem-se ao coeficiente de parentesco entre os indivíduos i e j .

Para composição da matriz A, Henderson (1976) descreve os seguintes passos:

- Codificar os indivíduos de 1 a N, ordenados de forma que os genitores precedam suas progêniess;

- Se ambos os genitores (s e d) do indivíduo i são conhecidos, fazer:

$$a_{ii} = 1 + 0,5 (a_{sd})$$

$$a_{ji} = a_{ij} = 0,5 (a_{js} + a_{jd}), j = 1 \text{ a } (i-1);$$

- Se somente um genitor (s) é conhecido e supostamente não aparentado com a fêmea com que cruzou, fazer:

$$a_{ii} = 1$$

$$a_{ji} = a_{ij} = 0,5 (a_{js}), j = 1 \text{ a } (i-1)$$

- Se ambos os genitores são desconhecidos e supostamente não aparentados, fazer:

$$a_{ii} = 1$$

$$a_{ji} = a_{ij} = 0, j = 1 \text{ a } (i-1)$$

Considere o seguinte pedigree:

Indivíduos	Pai	Mãe
4	1	Desconhecida
5	3	2
6	1	2
7	4	5
8	3	6

Para este pedigree, tem-se a seguinte matriz de parentesco:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0,5 & 0 & 0,5 & 0,25 & 0,25 \\ 0 & 1 & 0 & 0 & 0,5 & 0,5 & 0,25 & 0,25 \\ 0 & 0 & 1 & 0 & 0,5 & 0 & 0,25 & 0,5 \\ 0,5 & 0 & 0 & 1 & 0 & 0,25 & 0,5 & 0,125 \\ 0 & 0,5 & 0 & 0 & 1 & 0,25 & 0,5 & 0,375 \\ 0 & 0 & 0,5 & 0 & 0 & 1 & 0,25 & 0,5 \\ 0 & 0 & 0 & 0,5 & 0 & 0 & 1 & 0,25 \\ 0 & 0 & 0 & 0 & 0,5 & 0 & 0 & 1 \end{bmatrix}$$

Simétrica

Os elementos foram assim obtidos:

$$\begin{aligned} a_{11} &= 1 \\ a_{22} &= 1 \\ a_{33} &= 1 \\ a_{44} &= 1 \\ a_{55} &= 1 + 0 = 1 \\ a_{66} &= 1 + 0 = 1 \\ a_{77} &= 1 + 0 = 1 \\ a_{88} &= 1 + 0 = 1 \end{aligned}$$

$$\begin{aligned} a_{12} &= 0 \\ a_{13} &= 0 \\ a_{14} &= 0,5 (a_{js}) = 0,5 (a_{11}) = 0,5 \\ a_{15} &= 0,5 (a_{js} + a_{jd}) = 0,5 (a_{13} + a_{12}) = 0,5 (0) = 0 \\ a_{16} &= 0,5 (a_{11} + a_{12}) = 0,5 (1 + 0) = 0,5 \\ a_{17} &= 0,5 (a_{14} + a_{15}) = 0,5 (0,5 + 0) = 0,25 \\ a_{18} &= 0,5 (a_{13} + a_{16}) = 0,5 (0 + 0,5) = 0,25 \end{aligned}$$

Continua...

Continuação.

$a_{23} = 0$	$a_{34} = 0,5 (a_{31}) = 0$
$a_{24} = 0,5 (a_{21}) = 0$	$a_{35} = 0,5 (a_{33} + a_{32}) = 0,5$
$a_{25} = 0,5 (a_{23} + a_{22}) = 0,5 (0 + 1) = 0,5$	$a_{36} = 0,5 (a_{31} + a_{32}) = 0$
$a_{26} = 0,5 (a_{21} + a_{22}) = 0,5 (1) = 0,5$	$a_{37} = 0,5 (a_{34} + a_{35}) = 0,25$
$a_{27} = 0,5 (a_{24} + a_{25}) = 0,5 (0,5) = 0,25$	$a_{38} = 0,5 (a_{33} + a_{36}) = 0,5$
$a_{28} = 0,5 (a_{23} + a_{26}) = 0,5 (0,5) = 0,25$	
$a_{45} = 0,5 (a_{43} + a_{42}) = 0$	$a_{56} = 0,5 (a_{51} + a_{52}) = 0,25$
$a_{46} = 0,5 (a_{41} + a_{42}) = 0,25$	$a_{57} = 0,5 (a_{54} + a_{55}) = 0,5$
$a_{47} = 0,5 (a_{44} + a_{45}) = 0,5 (0,25 + 0,25) = 0,25$	$a_{58} = 0,5 (a_{53} + a_{56}) = 0,5(0,5 + 0,25) = 0,375$
$a_{48} = 0,5 (a_{43} + a_{46}) = 0,5 (0,25) = 0,125$	
$a_{67} = 0,5 (a_{64} + a_{65}) = 0,5 (0,25 + 0,25) = 0,25$	$a_{78} = 0,5 (a_{73} + a_{76}) = 0,5 (0,25 + 0,25) = 0,25$
$a_{68} = 0,5 (a_{63} + a_{65}) = 0,5$	

Como a matriz A é de elevada ordem (NXN), Henderson (1976) estabeleceu regras para escrever diretamente A^{-1} , sem a necessidade de inversão. Com genitores não endógamos, os seguintes passos devem ser adotados, partindo de uma matriz NXN, completamente zerada:

- Codificar os indivíduos de 1 a N, ordenados de forma que os genitores precedam suas progêniens.
- Se ambos os genitores (s e d) do indivíduo são conhecidos, adicionar:
 - 2 ao elemento a^{ii} .
 - 1 aos elementos a^{is} , a^{si} , a^{id} e a^{di} .
 - $\frac{1}{2}$ aos elementos a^{ss} , a^{sd} , a^{ds} e a^{dd} .
- Se apenas um dos genitores (s) do indivíduo i for conhecido, adicionar:
 - $\frac{4}{3}$ ao elemento a^{ii}
 - $-\frac{2}{3}$ aos elementos a^{is} e a^{si}
 - $\frac{1}{3}$ ao elemento a^{ss} .
- Se ambos os genitores do indivíduo i forem desconhecidos, adicionar 1 ao elemento a^{ii} .

Para o presente exemplo, tem-se:

$$A^{-1} = \begin{bmatrix} 1+1/3+1/2 & 1/2 & 0 & -2/3 & 0 & -1 & 0 & 0 \\ 1+1/2+1/2 & 1/2 & 0 & -1 & -1 & 0 & 0 & 0 \\ & & 1+1/2+1/2 & 0 & -1 & 1/2 & 0 & -1 \\ & & & 4/3+1/2 & 1/2 & 0 & -1 & 0 \\ & & & & 2+1/2 & 0 & -1 & 0 \\ & & & & & 2+1/2 & 0 & -1 \\ & & & & & & 2 & 0 \\ & & & & & & & 2 \end{bmatrix}$$

Simétrica

e, portanto:

$$A^{-1} = \begin{bmatrix} 1,833 & 0,500 & 0,000 & -0,667 & 0,000 & -1,000 & 0,000 & 0,000 \\ & 2,000 & 0,500 & 0,000 & -1,000 & -1,000 & 0,000 & 0,000 \\ & & 2,000 & 0,000 & -1,000 & 0,500 & 0,000 & -1,000 \\ & & & 1,833 & 0,500 & 0,000 & -1,000 & 0,000 \\ & & & & 2,500 & 0,000 & -1,000 & 0,000 \\ & & & & & 2,500 & 0,000 & -1,000 \\ & & & & & & 2,000 & 0,000 \\ & & & & & & & 2,000 \end{bmatrix}$$

Simétrica

Em determinadas situações, modificações na computação de A^{-1} são necessárias. Por exemplo, quando os genitores masculinos não são conhecidos com certeza, mas o grupo de k genitores masculinos potenciais é conhecido, pode-se concluir que a probabilidade de um específico genitor ser o pai é $1/k$ e pode-se computar o parentesco entre cada genitor e as progêniens como sendo $1/(2k)$. Outra modificação na computação de A^{-1} refere-se ao caso em que a variância genética aditiva difere entre grupos de indivíduos ou populações (Lynch & Walsh, 1997).

7.2.9.2. Matriz de parentesco de dominância

Os fundamentos estatísticos do uso da metodologia de modelos mistos para predição de efeitos aditivos e não aditivos (em ausência de endogamia) simultaneamente foram relatados por Henderson (1985). Dentro os efeitos não aditivos, os de dominância são mais relevantes que os epistáticos, nas populações alógamas.

Os coeficientes de parentesco de dominância entre indivíduos representam a probabilidade de que os mesmos apresentem genótipos idênticos em um mesmo loco. A matriz de parentesco de dominância, D , é constituída por coeficientes de parentesco de dominância e é essencial na predição dos efeitos de dominância pela metodologia de modelos mistos.

O parentesco de dominância entre um indivíduo X com genitores s e d e um indivíduo Y com genitores m e f , em uma população não endógama, pode ser calculado por (Cockerham, 1954):

$$\mu_{XY} = 0,25 (r_{sm} r_{df} + r_{sd} r_{mf}), \text{ em que:}$$

r_{sm} = representa o parentesco genético aditivo entre s e m .

Por exemplo, para dois indivíduos irmãos germanos, com pais não aparentados, tem-se $\mu_{XY} = 0,25 [(1) (1) + (0) (0)] = 0,25$. Assim, a matriz D pode ser gerada a partir do parentesco genético aditivo.

Para D de elevada ordem, a obtenção de D^{-1} pode se tornar impraticável. Para contornar este problema, Henderson (1985) propôs a utilização do seguinte sistema de equações de modelo misto, o qual exclui os efeitos de dominância, que são obtidos posteriormente.

$$\begin{bmatrix} X'X & X'ZK \\ Z'X & Z'ZK + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ em que:}$$

$$K = (I + DA^{-1} \sigma_d^2 / \sigma_a^2)$$

A solução para os efeitos de dominância é obtida por $\hat{d} = DA^{-1} \hat{a} \sigma_d^2 / \sigma_a^2$. Este procedimento requer a computação apenas de D e A^{-1} , evitando a obtenção de D^{-1} .

7.2.10. Acurácia das avaliações genéticas e número efetivo de informações

A matriz dos coeficientes das equações de modelo misto equivale a

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \text{ e a inversa generalizada de } C$$

é igual a $C^{-1} = \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$. O estimador da variância do erro de predição (PEV) dos valores genéticos é dado por:

$$PEV = Var(a - \hat{a}) = C^{22} \sigma_e^2 = (1 - r_{\hat{a}a}^2) \sigma_a^2, \text{ em que } r_{\hat{a}a} \text{ é a acurácia da predição dos valores genéticos.}$$

Assim, a acurácia pode ser estimada por: $r_{\hat{a}a} = [1 - PEV / \sigma_a^2]^{1/2}$.

Especificamente para um indivíduo i, tem-se:

$$PEV_i = d_i \sigma_e^2 = (1 - r_{\hat{a}_ia_i}^2) \sigma_a^2$$

$$r_{\hat{a}_ia_i} = (1 - d_i \sigma_e^2 / \sigma_a^2)^{1/2} = (1 - d_i \lambda)^{1/2}, \text{ em que:}$$

d_i : i-ésimo elemento da diagonal de C^{22} .

$$\lambda = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}.$$

O desvio padrão do erro de predição do valor genético do indivíduo i é dado por: $SEP_i = [d_i \sigma_e^2]^{1/2} = [(1 - r_{\hat{a}_ia_i}^2) \sigma_a^2]^{1/2}$.

É importante relatar que para o caso em que $R \neq I$ σ_e^2 e as equações de modelo misto não são fatoradas em termos de R^{-1} , $PEV = C^{22}$.

A acurácia e a PEV podem também ser expressas em termos do número efetivo (n_e) de informações do indivíduo, pelas expressões:

$$r_{\hat{a}_ia_i}^2 = \frac{n_e}{n_e + \lambda}$$

$$PEV = \frac{\lambda}{n_e + \lambda} \sigma_a^2$$

O número efetivo de informações reflete o número de informações disponíveis para a predição, após a correção para os efeitos fixos e outros efeitos aleatórios. Para um dado genitor, o número efetivo de informações pode ser computado por:

$$n_e = \sum_{i=1}^b \left(n_i - \frac{n_i}{n_b} \right), \text{ em que:}$$

n_i : número de descendentes do genitor no bloco i (ou no nível do efeito fixo).

n_b : número total de observações no bloco i .

Verifica-se, assim, que, à medida que o número de observações em cada nível do efeito fixo aumenta, aumenta também a acurácia.

De maneira reversa, o n_e pode ser obtido a partir da PEV por:

$$n_e = \left(\frac{1}{PEV_i} - \frac{1}{\sigma_a^2} \right) \sigma_e^2$$

7.3. BLUP Individual e Índice Multiefeitos com Uma Avaliação por Indivíduo

O modelo estatístico associado à avaliação de p progênies, em b blocos com n plantas por parcela equivale a:

$$Y_{ijk} = \mu + p_i + b_j + e_{ij} + d_{ijk}, \text{ em que:}$$

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$.

p_i : efeito da progénie ou família i , aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$.

b_j : efeito de bloco ou repetição j , fixo, $E(b_j) = b_j$ e $E(b_j^2) = V_b$.

e_{ij} : efeito da parcela ij , aleatório, $E(e_{ij}) = 0$ e $E(e_{ij}^2) = \sigma_e^2$.

d_{ijk} : efeito do indivíduo k dentro da parcela ij , aleatório, $E(d_{ijk}) = 0$ e $E(d_{ijk}^2) = \sigma_d^2$.

7.3.1. Seleção por valores genéticos aditivos (propagação sexuada)

Nesta situação, o método de seleção índice multiefeitos (que considera todos os efeitos do modelo estatístico, exceto o de bloco, que não afeta o mérito genético), é dado por Resende & Higa (1994a):

$$I_1 = b_1(Y_{ijk} - \bar{Y}_{ij.}) + b_2(\bar{Y}_{i..} - \bar{Y}_{...}) + b_3(\bar{Y}_{ij.} - \bar{Y}_{i..} + \bar{Y}_{...}), \text{ em que:}$$

$$b_1 = h_{\delta}^2 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_{\delta}^2} = \frac{(1-\rho_a) h_a^2}{1-t_1};$$

$$b_2 = h_m^2 = \frac{\{[1+(nb-1) \rho_a]/(nb)\} \sigma_a^2}{\sigma_p^2 + \sigma_e^2/b + \sigma_{\delta}^2/(nb)} = \frac{[1+(nb-1) \rho_a] h_a^2}{1+(n-1) t_1 + (b-1) nt_2};$$

$$b_3 = h_p^2 = \frac{[(1-\rho_a)/n] \sigma_a^2}{\sigma_e^2 + \sigma_{\delta}^2/n} = \frac{(1-\rho_a) h_a^2}{1+(n-1) t_1 - nt_2}.$$

em que:

h_{δ}^2 , h_m^2 e h_p^2 : herdabilidades dos efeitos de indivíduos dentro de parcela, de progênies e de parcela, respectivamente.

σ_a^2 , σ_p^2 , σ_e^2 e σ_{δ}^2 : componentes de variância genética aditiva, entre progênies, entre parcelas e dentro de parcelas, respectivamente.

$\sigma_y^2 = \sigma_p^2 + \sigma_e^2 + \sigma_{\delta}^2$: variância fenotípica ao nível de indivíduos.

$h_a^2 = \sigma_a^2 / \sigma_y^2$: herdabilidade individual no sentido restrito.

$t_2 = \rho_a h_a^2$: correlação intraclasse entre os indivíduos de uma mesma família em diferentes blocos.

$t_1 = t_2 + c^2$: correlação intraclasse entre os indivíduos de uma mesma parcela.

$c^2 = \frac{\sigma_e^2}{\sigma_y^2}$: correlação entre os indivíduos da parcela devida ao seu ambiente comum.

ρ_a : correlação genética intraclasse entre indivíduos da progénie ($\rho_a = 0,25$ para progênies de meios-irmãos e $\rho_a = 0,5$ para progênies de irmãos germanos).

$\bar{Y}_{..}$, $\bar{Y}_{.j}$, $\bar{Y}_{i..}$ e $\bar{Y}_{ij..}$: médias geral, de bloco, de progénie e de parcela, respectivamente.

No contexto dos modelos lineares mistos, tem-se o modelo:

$y = X\beta + Za + Wp + e$, em que:

y , β , a , p e e : vetores de dados, de efeitos fixos (média geral e efeitos de bloco), de efeitos genéticos aditivos (aleatório), de efeitos de parcelas (aleatório) e de erros aleatórios, respectivamente.

X , Z e W : matrizes de incidência para β , a e p , respectivamente.

Derivando-se o índice multiefeitos com base neste modelo misto (no qual o efeito de bloco é considerado como fixo), tem-se:

$$I_2 = b_4 C_1 + b_5 C_2 + b_6 C_3$$

$$= b_4[Y_{ijK} - E(Y_{ijK})] + b_5[\bar{Y}_{i..} - E(\bar{Y}_{i..})] + b_6[\bar{Y}_{ij.} - E(\bar{Y}_{ij.})]$$

$$= b_4(Y_{ijK} - \bar{Y}_{j..}) + b_5(\bar{Y}_{i..} - \bar{Y}_{...}) + b_6(\bar{Y}_{ij.} - \bar{Y}_{j..}), \text{ pois}$$

$$E(Y_{ijK}) = \mu + b_j = \bar{Y}_{j..};$$

$$E(\bar{Y}_{i..}) = \mu = \bar{Y}_{...};$$

$$E(\bar{Y}_{ij.}) = \mu + b_j = \bar{Y}_{j..}.$$

Como existem covariâncias entre os componentes (C_i) deste índice, os valores de b devem ser determinados por:

$b = P^{-1}G$, em que:

$$b = \begin{bmatrix} b_4 \\ b_5 \\ b_6 \end{bmatrix}; \quad P = \begin{bmatrix} \sigma_{c_1}^2 & \sigma_{c_1 c_2} & \sigma_{c_1 c_3} \\ \sigma_{c_1 c_2} & \sigma_{c_2}^2 & \sigma_{c_2 c_3} \\ Simétrica & \sigma_{c_2 c_3} & \sigma_{c_3}^2 \end{bmatrix}; \quad G = \begin{bmatrix} \sigma_{c_1 a} \\ \sigma_{c_2 a} \\ \sigma_{c_3 a} \end{bmatrix}$$

$$\sigma_{C_1}^2 = Var(Y_{ijK} - \bar{Y}_{j..}) = \frac{p-1}{p}(\sigma_p^2 + \sigma_e^2) + \frac{np-1}{np} \sigma_d^2$$

$$\sigma_{C_2}^2 = Var(\bar{Y}_{i..} - \bar{Y}_{...}) = \frac{p-1}{p}(\sigma_p^2 + \sigma_e^2 / b + \sigma_d^2 / (nb))$$

$$\sigma_{C_3}^2 = Var(\bar{Y}_{ij.} - \bar{Y}_{j..}) = \frac{p-1}{p}(\sigma_p^2 + \sigma_e^2 + \sigma_d^2 / n)$$

Derivando-se as covariâncias da matriz P , obtém-se as igualdades:

$$\sigma_{C_1 C_2} = \text{cov}(C_1, C_2) = \sigma_{C_2}^2$$

$$\sigma_{C_1 C_3} = \text{cov}(C_1, C_3) = \sigma_{C_3}^2$$

$$\sigma_{C_2 C_3} = \text{cov}(C_2, C_3) = \sigma_{C_2}^2$$

$$\sigma_{C_1 a} = \text{cov}(C_1, a) = \frac{(np-1)-(n-1)}{np} \rho_a \sigma_a^2$$

$$\sigma_{C_2 a} = \text{cov}(C_2, a) = \frac{p-1}{p} \frac{1+(nb-1)}{nb} \rho_a \sigma_a^2$$

$$\sigma_{C_3 a} = \text{cov}(C_3, a) = \frac{p-1}{np} \sigma_a^2 + \frac{p(n-1)-(n-1)}{np} \rho_a \sigma_a^2$$

As equações de modelo misto para predição dos valores genéticos pelo procedimento BLUP sob modelo individual (que devem fornecer os mesmos resultados que o índice I_2 em situações de dados balanceados, já que é a versão matricial do mesmo) equivalem a (Resende, 1999b; Resende & Fernandes, 1999):

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

$$\lambda_1 = \frac{\sigma_{rd}^2}{\sigma_a^2} = \frac{1-h_a^2-c^2}{h_a^2}; \quad \lambda_2 = \frac{\sigma_{rd}^2}{\sigma_e^2} = \frac{1-h_a^2-c^2}{c^2}$$

$\sigma_{rd}^2 = \sigma_\delta^2 - (1-\rho_a) \sigma_a^2$: variância residual (ambiental + genética não aditiva).

A : matriz de parentesco genético aditivo entre os indivíduos.

Para a definição dos elementos de β , recomenda-se a adoção de um modelo de médias, de forma que β contenha as médias de blocos e não os seus efeitos. Assim, a matriz dos coeficientes terá inversa única.

A composição dos efeitos genéticos preditos pelo BLUP é dada por:

$$\hat{a}_i = (\hat{a}_{pi} + \hat{a}_{mi})/2 + h_{\delta}^2(Y_{ijK} - \hat{\beta}_{.j.} - (\hat{a}_p + \hat{a}_m)/2 - \hat{p}_{ij}), \text{ em que:}$$

\hat{a}_{pi} , \hat{a}_{mi} e \hat{p}_{ij} : efeitos genéticos preditos do pai e da mãe do indivíduo e efeito de parcela predito, respectivamente.

$$\hat{p}_{ij} = h_{pp}^2(\bar{Y}_{ij.} - \bar{\beta}_{.j.} - \bar{a}).$$

$$h_{pp}^2 = \sigma_e^2 / \{\sigma_e^2 + [\sigma_\delta^2 - (1-\rho_a) \sigma_a^2] / n_{ij}\}.$$

\bar{a} - média dos efeitos genéticos preditos dos indivíduos da parcela.

$$\hat{a}_{pi} = \hat{a}_{mi} = h_{mm}^2(\bar{Y}_{i..} - \bar{Y}_{...}) \quad (\text{no caso de progêneres de meios-irmãos } \hat{a}_{pi} = 0)$$

$$h_{mm}^2 = [(1/2) \sigma_a^2] / [\sigma_p^2 + \sigma_e^2 / b_i + \sigma_\delta^2 / \sum n_i] = \frac{(1/2) \sum n_i h_a^2}{1 + (\sum n_i / \sum b_i - 1) t_1 + (\sum b_i - 1) (\sum n_i / \sum b_i) t_2}$$

$$\text{Para o caso balanceado, } h_{mm}^2 = \frac{(1/2)nb h_a^2}{1 + (n-1) t_1 + (b-1) nt_2}.$$

A acurácia da predição de valores genéticos pelo índice multiefeitos é dada por:

$$\hat{r}_{la} = [\alpha_1 \hat{b}_1 + \alpha_2 \hat{b}_2 + \alpha_3 \hat{b}_3]^{1/2}$$

$$\text{em que: } \alpha_1 = (1-\rho_a) \frac{n-1}{n}; \quad \alpha_2 = \frac{1+(nb-1)\rho_a}{nb} \frac{p-1}{p}$$

$$\alpha_3 = \frac{(1-\rho_a)}{n} \frac{p-1}{p} \frac{b-1}{b}$$

Para a predição dos efeitos aditivos pelo procedimento BLP (melhor predição linear, conforme Henderson, 1984), os seguintes estimadores associados ao índice I_1 devem ser empregados:

$$b_1 = \frac{(1-\rho_a) h_a^2}{1-t_1}$$

$$b_2^* = \frac{[1+(\sum n_i - 1) \rho_a] h_a^2}{1 + (\sum n_i / \sum b_i - 1) t_1 + (\sum b_i - 1) (\sum n_i / \sum b_i) t_2}$$

$$b_3^* = \frac{(1-\rho_a) h_a^2}{1 + (n_{ij} - 1) t_1 - n_{ij} t_2}, \text{ em que:}$$

$\sum n_i$, b_i e n_{ij} : somatório do número de indivíduos na progênie i , número de blocos em que está presente a progênie i e número de plantas na parcela ij , respectivamente.

7.3.2. Seleção por valores genotípicos (propagação assexuada)

A ordenação dos indivíduos por seus valores genotípicos preditos (Resende & Higa, 1994b), visando à seleção e propagação vegetativa, pode ser realizada pelo índice multifeitos clonal, dado por:

$$I_3 = b_7(Y_{ijK} - \bar{Y}_{ij.}) + b_8(\bar{Y}_{i..} - \bar{Y}_{...}) + b_9(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}), \text{ em que:}$$

$$\begin{aligned} b_7 &= h_{dc}^2 = \frac{(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2}{\sigma_\delta^2} = \\ &= \frac{(1-\rho_a) h^2 + (1-\rho_d)(h_g^2 - h_a^2)}{1-t_1} \end{aligned}$$

$$\begin{aligned} b_8 &= h_{mc}^2 = \frac{[1+(nb-1) \rho_a] \sigma_a^2 + 1 + (nb-1) \rho_d] \sigma_d^2 / (nb)}{\sigma_p^2 + \sigma_e^2 / b + \sigma_\delta^2 / (nb)} = \\ &= \frac{[1+(nb-1) \rho_a] h_a^2 + [1+(nb-1) \rho_d](h_g^2 - h_a^2)}{1 + (n-1) t_1 + (b-1) nt_2} \end{aligned}$$

$$\begin{aligned} b_9 &= h_{pc}^2 = \frac{[(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2] / n}{\sigma_e^2 + \sigma_\delta^2 / n} = \\ &= \frac{(1-\rho_a) h_a^2 + (1-\rho_d)(h_g^2 - h_a^2)}{1 + (n-1) t_1 - nt_2} \end{aligned}$$

h_{dc}^2 , h_{mc}^2 e h_{pc}^2 : herdabilidades no sentido amplo dos efeitos de indivíduos dentro de parcela, de progênie e de parcela, respectivamente.

h_g^2 : herdabilidade individual no sentido amplo.

ρ_d : coeficiente de parentesco de dominância.

No contexto dos modelos lineares mistos, tem-se o modelo:

$$y = X\beta + Za + Zd + Wp + e, \text{ em que } d \text{ é o vetor de efeitos de dominância.}$$

As equações de modelo misto para a predição dos efeitos aditivos e de dominância pelo procedimento BLUP sob modelo individual equivalem a (Resende, 1999b; Resende & Fernandes, 1999):

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{d} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_3 & Z'Z & Z'W \\ Z'X & Z'Z & Z'Z + D^{-1}\lambda_4 & Z'W \\ W'X & W'Z & W'Z & W'W + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ Z'y \\ W'y \end{bmatrix} \quad (2), \text{ em que:}$$

$$\lambda_3 = \frac{\sigma_{ad}^2}{\sigma_a^2} = \frac{1-h_g^2 - c^2}{h_a^2}; \quad \lambda_4 = \frac{\sigma_{ad}^2}{\sigma_d^2} = \frac{1-h_g^2 - c^2}{(h_g^2 - h_a^2)}; \quad \lambda_5 = \frac{\sigma_{ad}^2}{\sigma_e^2} = \frac{1-h_g^2 - c^2}{c^2}$$

$\sigma_{ad}^2 = \sigma_\delta^2 - (1-\rho_a)$ $\sigma_a^2 - \sigma_d^2$: variância ambiental dentro de parcela.

D : matriz de parentesco de dominância entre indivíduos.

$\sigma_d^2 = (h_g^2 - h_a^2)$ σ_y^2 : variância genética de dominância.

$\sigma_y^2 = \sigma_p^2 + \sigma_e^2 + \sigma_\delta^2$: variância fenotípica individual.

h_g^2 : herdabilidade individual no sentido amplo.

h_a^2 : herdabilidade individual no sentido restrito.

Os valores genotípicos preditos dos indivíduos são dados por: $\hat{g} = \hat{a} + \hat{d}$.

Sendo $Var(g) = A \sigma_a^2 + D \sigma_d^2 = G$, os valores genotípicos podem ser preditos diretamente por:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + G^{-1}\sigma_{ad}^2 & Z'W \\ W'X & W'Z & W'W + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix} \quad (3)$$

A composição dos efeitos de dominância preditos pelo BLUP é dada por:

$$d_i = (Y_{ijK} - \hat{\beta}_{.j.} - \hat{a}_i - \hat{p}_i) h_{dom}^2, \text{ em que } h_{dom}^2 = \frac{\sigma_d^2}{\sigma_d^2 + \sigma_{ad}^2}$$

7.3.3. Teste clonal no delineamento de blocos ao acaso com várias plantas por parcela

O modelo estatístico é o mesmo apresentado no item 7.3.1, apenas substituindo p_i por c_i , onde c_i é o efeito do clone i .

No contexto dos modelos lineares mistos, tem-se:

$y = X\beta + Z_c c + Wp + e$, em que:

c e Z_c – vetor dos efeitos de clones (aleatório) e sua matriz de incidência, respectivamente.

As equações de modelo misto são exatamente conforme apresentado em (3) e as predições dos valores genotípicos de cada clone são dadas por (considerando clones não aparentados):

$$\hat{g}_i = h_c^2 (\bar{Y}_{..} - \bar{Y}_{..}) \text{ em que } h_c^2 = (\sigma_a^2 + \sigma_d^2) / [(\sigma_a^2 + \sigma_d^2) + \sigma_e^2 / b_i + \sigma_{ad}^2 / \sum n_i].$$

Para o caso balanceado $H_c^2 = \frac{nb h_g^2}{1 + (n-1) (h_g^2 + c^2) + (b-1) nh_g^2}$: herdabilidade associada a médias de clone.

7.3.4. Teste de progênie no delineamento de blocos ao acaso com uma planta por parcela: seleção por valores genéticos aditivos

Neste caso, o modelo estatístico se resume a:

$$Y_{ij} = \mu + p_i + b_j + e_{ij}$$

e o índice multiefeitos para valores genéticos aditivos equivale a (Resende & Higa, 1994b):

$$I_4 = b_{10}(\bar{Y}_{ij} - \bar{Y}_{..} - \bar{Y}_{..j} + \bar{Y}_{..}) + b_{11}(\bar{Y}_{..} - \bar{Y}_{..}), \text{ em que}$$

$$b_{10} = h_{d\mu}^2 = \frac{(1 - \rho_a) \sigma_a^2}{\sigma_e^2}; \quad b_{11} = h_{m\mu}^2 = \frac{\{[1 + (b-1) \rho_a]/b\} \sigma_a^2}{\sigma_p^2 + \sigma_e^2 / b}$$

Em termos matriciais, tem-se:

$$y = X\beta + Za + e,$$

e as equações de modelo misto para predição dos valores genéticos aditivos e estimação das médias de blocos são:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda_6 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix} (4), \text{ em que}$$

$$\lambda_6 = \frac{\sigma_r^2}{\sigma_a^2} = \frac{1 - h_a^2}{h_a^2}; \quad \sigma_r^2 = \sigma_e^{2*} - (1 - \rho_a) \sigma_a^2$$

σ_e^{2*} : variância entre parcelas estimada a partir de delineamento com uma planta por parcela.

O índice multiefeitos apresentado é equivalente à seleção combinada utilizando os valores individuais ajustados para os efeitos de blocos, conforme demonstrado por Resende (1991). Definindo os efeitos de blocos como fixos, a correção dos valores fenotípicos individuais pelas equações de modelo misto é realizada exatamente por meio do ajuste para os efeitos de blocos. Resende (1991) demonstrou também que os coeficientes (b) da seleção combinada nes-

se caso são equivalentes para os delineamentos inteiramente casualizados com uma planta por parcela e blocos ao acaso com uma planta por parcela, com ajuste para os efeitos de blocos. No caso do delineamento inteiramente casualizado com uma planta por parcela, o modelo estatístico equivale a $Y_{ij} = \mu + p_i + e_{ij}$ (modelo muito utilizado no melhoramento animal) e o índice de seleção ou seleção combinada é dado por:

$I_5 = b_{12}(Y_{ij} - \bar{Y}_{..}) + b_{13}(\bar{Y}_{..} - \bar{Y}_{..})$, em que $b_{12} = b_{10}$ e $b_{13} = b_{11}$. Dessa forma, as equações de modelo misto com um único efeito aleatório, desenvolvidas e muito aplicadas no melhoramento animal são adequadas ao melhoramento florestal, apenas para delineamentos experimentais (em blocos ou não) com uma planta por parcela.

7.3.5. Exemplo prático

Foram utilizados dados referentes à variável diâmetro à altura do peito, avaliada em um teste de 33 progêneres de meios-irmãos de *Eucalyptus grandis* instaladas no delineamento de blocos ao acaso com 6 repetições e 6 plantas por parcela. A partir destes dados, foram estimados os seguintes componentes de variância:

$$\hat{\sigma}_p^2 = 0,2722; \hat{\sigma}_e^2 = 0,5187; \hat{\sigma}_\delta^2 = 5,8672; \hat{\sigma}_y^2 = 6,6581; \hat{\sigma}_a^2 = 1,0888 \\ \hat{h}_a^2 = 0,1635; \hat{\sigma}_{rd}^2 = 5,0506$$

Para predição dos valores genotípicos, considerou-se: $\sigma_d^2 / \sigma_a^2 = 0,5$ e, portanto, $\hat{h}_g^2 = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_d^2}{\hat{\sigma}_y^2} = 0,2453$ e $\hat{\sigma}_d^2 = 0,5444$.

Assim, as seguintes relações entre componentes de variância foram utilizadas:

$$\lambda_1 = 4,6387; \lambda_2 = 9,7370; \lambda_3 = 4,1387; \lambda_4 = 8,2774; \lambda_5 = 8,6875.$$

Considerou-se $\hat{\sigma}_e^{2*} = \hat{\sigma}_e^2 + \hat{\sigma}_\delta^2$ e, portanto, $\lambda_6 = 5,1151$.

Para aplicação das metodologias, considerou-se o seguinte subconjunto de dados, extraído do arquivo original referente ao experimento em questão:

Indivíduo	Bloco	Progênie	Árvore	Diâmetro (cm)
4	1	1	1	9,87
5	1	1	2	14,48
6	1	2	1	8,91
7	1	2	2	14,64
8	1	3	1	9,55

Continua...

Continuação.

Indivíduo	Bloco	Progênie	Árvore	Diâmetro (cm)
9	1	3	2	7,96
10	2	1	1	16,07
11	2	1	2	14,01
12	2	2	1	7,96
13	2	2	2	21,17
14	2	3	1	10,19
15	2	3	2	9,23

A matriz de parentesco genético aditivo (A), com a inclusão das 3 genitoras equivale a:

$$\begin{bmatrix} 1.0000 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0 & 0.5000 & 0.5000 \\ 0.5000 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5000 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 \\ 0.5000 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5000 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 \end{bmatrix}$$

A matriz de parentesco de dominância (D) é uma matriz identidade, visto que os indivíduos e matrizes, neste caso (progênie de meios-irmãos), não possuem parentesco de dominância.

Na seqüência são apresentados os resultados da aplicação prática.

Teste de progênie no delineamento de blocos ao acaso com várias plantas por parcela

Seleção por valores genéticos aditivos

Os resultados referentes aos índices I_1 e I_2 e BLUP(1), bem como seus ponderadores são apresentados na Tabela 7.1.

Tabela 7.1. Coeficientes de ponderação (b_i e h_i^2) das diferentes informações fenotípicas e valores genéticos preditos por diferentes procedimentos.

Ponderado-res do índice I_1	Ponderado-res do índice I_2	Ponderadores associados ao BLUP(1)	Indivíduos	Efeitos aditivos preditos pelo índice I_1	Efeitos aditivos preditos pelo índice I_2	Efeitos aditivos preditos pelo BLUP(1)
$b_1=0,1392$	$b_4=0,1392$	$h_{\delta}^2 = 0,1392$	1	-	-	0,4370
$b_2=0,2384$	$b_5=0,1201$	$h_{pp}^2 = 0,1704$	2	-	-	0,3178
$b_3=0,1183$	$b_6=0,0209$	$h_{mm}^2 = 0,2724$	3	-	-	-0,7548
			4	0,0224	0,0224	0,0224
			5	0,6641	0,6641	0,6641
			6	-0,1554	-0,1554	-0,1554
			7	0,6422	0,6422	0,6422
			8	-0,4760	-0,4760	-0,4760
			9	-0,6973	-0,6973	-0,6973
			10	0,5649	0,5649	0,5649
			11	0,2782	0,2782	0,2782
			12	-0,6065	-0,6065	-0,6065
			13	1,2321	1,2321	1,2321
			14	-0,6675	-0,6675	-0,6675
			15	-0,8011	-0,8011	-0,8011

Verifica-se que os três procedimentos, embora computacionalmente distintos (em termos de seus componentes e ponderadores) conduzem a resultados idênticos, para o caso de dados balanceados. I_2 é o índice de referência por constituir-se no índice ótimo derivado sob modelo misto, e o procedimento BLUP(1) é a versão matricial e generalizada do índice I_2 e, por esperança, os dois procedimentos deveriam realmente fornecer os mesmos resultados. A coincidência entre os índices I_1 e I_2 indica que os índices multiefeitos, do tipo I_1 , podem ser utilizados eficientemente para derivação de seus equivalentes BLUP, em termos de equações de modelo misto. Os índices multiefeitos tipo I_1 devem ser preferidos, pois são derivados com mais facilidade, visto que, neste caso, os componentes do índice ou efeitos do modelo estatístico não apresentam covariância. Logicamente, os três procedimentos conduzem (para o caso balanceado), em essência, aos mesmos ponderadores dos valores individuais (Y_{ijk}), das médias de famílias ($\bar{Y}_{i..}$), das médias de parcela ($\bar{Y}_{ij..}$), das médias de bloco ($\bar{Y}_{...j}$) e da média geral ($\bar{Y}_{...}$), sendo o índice geral dado por

$$I = h_{\delta}^2 Y_{ijk} + (h_m^2 - h_p^2) \bar{Y}_{i..} + (h_p^2 - h_{\delta}^2) \bar{Y}_{ij..} - h_p^2 \bar{Y}_{...j} + (h_m^2 - h_p^2) \bar{Y}_{...}, \text{ que no caso equiva-} \\ \text{le a}$$

$$I = 0,1392 Y_{ijk} + 0,1201 \bar{Y}_{i..} - 0,0209 \bar{Y}_{ij..} - 0,1183 \bar{Y}_{...j} - 0,1201 \bar{Y}_{...}.$$

Os resultados obtidos permitem afirmar também que, para o caso balanceado, o método índice multiefeitos é BLUP, ao passo que as seleções combinadas formuladas com base no desvio do valor individual em relação a média de parcela (Bueno Filho, 1992; Resende & Higa, 1994b) e no desvio do valor individual em relação à média do bloco (Pires et al., 1996) não são BLUP.

O procedimento BLUP (1) fornece, também, ao mesmo tempo, a predição dos valores genéticos dos indivíduos (\hat{a}), a estimativa dos efeitos fixos ($\hat{\beta}$) e a predição dos efeitos aleatórios de parcela (\hat{p}). No presente caso, foram estimadas as médias de blocos: $\hat{\beta}_1 = \bar{Y}_{1.} = 10,9017$; $\hat{\beta}_2 = \bar{Y}_{2.} = 13,1050$ e preditos os efeitos ambientais de parcela: $\hat{p}_1 = 0,1585$; $\hat{p}_2 = 0,1073$; $\hat{p}_3 = -0,2658$; $\hat{p}_4 = 0,2579$; $\hat{p}_5 = 0,1955$ e $\hat{p}_6 = -0,4534$. O BLUP prediz os valores genéticos também dos genitores (primeiros três elementos da Tabela 7.1, última coluna), fato que não é possível com o índice multiefeitos.

É lógico que o procedimento BLUP apresenta as suas vantagens no caso de dados desbalanceados, em que conduz a estimativas BLUE (melhor estimativa linear não viciada) dos efeitos fixos empregando-se o procedimento de mínimos quadrados generalizados e também a predições não viciadas dos valores genéticos e mínima variância do erro de predição, considerando também as diferentes quantidades de dados associadas aos diferentes candidatos à seleção. As propriedades favoráveis do procedimento BLUP em relação aos seus competidores são bem documentadas em literatura (Henderson et al., 1959; Bulmer, 1980; Gianola & Goffinet, 1982; Ronningen & Van Vleck, 1985; Van Vleck, 1993a). Todavia, em situações em que a distribuição normal multivariada não pode ser considerada, podem existir preditores com melhores propriedades de ordenamento do que o BLUP (Portnoy, 1982).

Seleção por valores genotípicos

Os resultados referentes ao índice I_3 e procedimentos BLUP (2) e BLUP (3), bem como seus ponderadores são apresentados na Tabela 7.2.

Tabela 7.2. Coeficientes de ponderação (b_i e h_i^2) das diferentes informações fenotípicas, efeitos aditivos, de dominância e efeitos genotípicos preditos por diferentes procedimentos.

Ponderadores do índice I_3	Ponderador associado ao Ponderador BLUP(2) associado ao	Indivíduos	Efeitos genotípicos preditos pelo índice I_3	Efeitos genotípicos preditos pelo procedimento BLUP(2)	Efeitos aditivos	Efeitos de dominância
						I_2 e BLUP(2)
$b_7=0,2320$	$BLUP(2)$	4	-0,1083	0,1083	0,0224	-0,1307
$b_8=0,3065$		5	0,9611	0,9611	0,6641	0,2970
$b_9=0,1971$		6	-0,3649	-0,3649	-0,1553	-0,2095
		7	0,9643	0,9643	0,6422	0,3222
		8	-0,5417	-0,5417	-0,4760	-0,0657
		9	-0,9106	-0,9106	-0,6973	-0,2133
		10	0,7958	0,7958	0,5649	0,2309
		11	0,3179	0,3179	0,2782	0,0398
		12	-1,1168	-1,1168	-0,6065	-0,5103
		13	1,9475	1,9475	1,2321	0,7154
		14	-0,8609	-0,8609	-0,6675	-0,1934
		15	-1,0836	-1,0836	-0,8011	-0,2825

Verifica-se que o índice multiefeitos e o BLUP conduziram a resultados idênticos, fato que confirma a veracidade das equações de modelo misto apresentadas. Os valores de $(\hat{\beta}_i \text{ e } \hat{p}_i)$ obtidos pelo BLUP (2) e BLUP (3) são idênticos aos apresentados no tópico anterior. Os valores genotípicos dos genitores, obtidos pelo BLUP (3) são idênticos aos seus valores genéticos aditivos preditos pelo BLUP (1), visto que seus efeitos de dominância (d) foram preditos como zero, uma vez que os genitores não possuíam observações próprias.

Comparando-se os pesos b_7 , b_8 e b_9 com b_1 , b_2 e b_3 , respectivamente, verifica-se que ocorreram aumentos consideráveis em seus valores, quando também foram incluídos os efeitos de dominância no objetivo da seleção.

Teste clonal no delineamento de blocos ao acaso com várias plantas por parcela

Considerando no subconjunto de dados as progénies como clones e indivíduos dentro de progénies como seus ramos, estimou-se $h_c^2 = 0,5410$, e as soluções das equações de modelo misto pelo procedimento BLUP(3) são apresentadas na Tabela 7.3.

Tabela 7.3. Soluções das equações de modelo misto pelo procedimento BLUP(3), considerando-se um teste clonal (3 clones) no delineamento de blocos (2 blocos) ao acaso com dois indivíduos por parcela.

Efeitos	Soluções
Efeitos fixos	
Média do bloco 1 = $\hat{\mu} + \hat{b}_1$	10,9017
Média do bloco 2 = $\hat{\mu} + \hat{b}_2$	13,1050
Efeitos aleatórios	
Efeitos genotípicos	
Clone 1 = \hat{g}_1	0,8678
Clone 2 = \hat{g}_2	0,6311
Clone 3 = \hat{g}_3	-1,4989
Efeitos de parcela	
\hat{p}_1	0,0759
\hat{p}_2	0,0453
\hat{p}_3	-0,1212
\hat{p}_4	0,1997
\hat{p}_5	0,1551
\hat{p}_6	-0,3548

Os resultados obtidos para os valores genotípicos equivalem ao que seria obtido considerando a análise convencional de teste clonal balanceado.

Teste de progênie no delineamento de blocos ao acaso com uma planta por parcela: seleção por valores genéticos aditivos

Os resultados referentes ao índice I_4 e procedimento BLUP(4) são apresentados na Tabela 7.4. Do conjunto de dados, considerou-se apenas a primeira planta de cada parcela.

Tabela 7.4. Coeficientes de ponderação (b_i) das diferentes informações fenotípicas e valores genéticos aditivos preditos por diferentes procedimentos.

Ponderadores do Índice I_4	Indivíduos	Efeitos aditivos preditos pelo índice I_4	Efeitos aditivos preditos pelo BLUP(4)	Médias de Blocos e efeitos aditivos dos genitores preditos pelo BLUP(4)
$b_{10}=0,1279$	4	0,2289	0,2289	$\hat{\beta}_1 = 9,4433$
$b_{11}=0,1964$	6	-0,2045	-0,2045	$\hat{\beta}_2 = 11,4067$
	8	-0,0244	-0,0244	$\hat{a}_1 = 0,3998$
	10	0,7707	0,7707	$\hat{a}_2 = -0,3126$
	12	-0,5771	-0,5771	$\hat{a}_3 = -0,0872$
	14	-0,1936	-0,1936	

Os resultados obtidos pelo índice I_4 e procedimento BLUP(4) foram também idênticos, conforme esperado.

7.3.6. Definição do bloco como principal efeito fixo associado às avaliações genéticas em plantas perenes

O vetor de efeitos fixos deve contemplar os efeitos ambientais identificáveis, e o principal efeito ambiental identificável nos experimentos com plantas perenes é o efeito de blocos, o qual praticamente não contribui (Resende & Higa, 1994) com informação genética a respeito dos indivíduos. O bloco é o principal estrato homogêneo para comparação de indivíduos, de forma que dentro de blocos os indivíduos são prontamente comparáveis de uma maneira não viciada. Para comparação de indivíduos através dos blocos, torna-se necessário ajustar os valores individuais para os efeitos de blocos, de maneira que as comparações sejam não viciadas. A forma de conseguir isto é tratar os efeitos de blocos como fixos, fato que torna as avaliações genéticas invariantes aos efeitos de blocos e não viciadas devido à associação (correlação) entre efeitos de blocos e nível genético dos indivíduos que nele se desenvolvem.

A obtenção de avaliações genéticas invariantes aos efeitos de repetições, ou seja, comparáveis de maneira não viciada nas diferentes repetições, depende da consideração do efeito de repetições como fixo e também de uma boa conectabilidade (capacidade para estimar diferenças ambientais entre repetições, conforme Weeks & Williams, 1964) entre as repetições, o que é conseguido procurando-se fazer a maioria das progênies estar representada em cada nível do efeito fixo.

A definição dos efeitos de blocos como fixos parece contraditória em relação ao nome do delineamento: blocos ao acaso. Entretanto, a rigor, o nome correto é “delineamento de tratamentos aleatorizados em blocos”, uma vez que os blocos podem ser considerados como fixos ou aleatórios, dependendo da situação experimental. Assim, os tratamentos é que são casualizados e não necessariamente os blocos.

De maneira genérica, os blocos, quando completos, devem ser tratados como efeitos fixos e, quando incompletos (como nos delineamentos em látice) devem ser considerados aleatórios. Os blocos incompletos retêm informações intertratamentos e devem ser considerados aleatórios, visando recuperar esta informações. Maiores detalhes sobre a definição de efeitos fixos e aleatórios são apresentados no tópico 7.8. Resende (2002) discute com detalhes a consideração de repetições como efeitos fixos ou aleatórios.

Considerações finais

(i) O método de seleção índice multiefeitos é BLUP para situação de dados balanceados. As duas formas de seleção combinada não são BLUP e podem ser consideradas índices imperfeitos, mesmo para o caso balanceado; (ii) para derivação do procedimento BLUP associado a diferentes estruturas experimentais, pode-se adotar a estratégia de derivação de seus correspondentes índices multiefeitos; (iii) as equações de modelo misto apresentadas para as situações específicas permitem uma acurada predição de valores genéticos e genotípicos em situações de dados desbalanceados, sendo prontamente aplicáveis e, portanto, de grande utilidade prática em melhoramento genético de espécies perenes.

7.4. BLUP Individual e Índice Multiefeitos com Medidas Repetidas

A utilização de medidas repetidas (no tempo) em cada indivíduo é prática rotineira no melhoramento genético de espécies perenes. A avaliação genética de indivíduos dessas espécies deve, então, basear-se em modelos que considerem o efeito adicional denominado efeito de ambiente permanente, bem como o parâmetro associado à correlação fenotípica entre medidas repetidas em um mesmo indivíduo, o qual é denominado repetibilidade. A seguir, são apresentados os desenvolvimentos realizados por Resende (1999b) e Resende et al. (2000a).

Considere o seguinte modelo estatístico associado à avaliação de p progênies em b blocos com n plantas por parcela e m medições por planta, no delineamento de blocos ao acaso:

$$\bar{Y}_{ijk.} = \mu + p_i + b_j + e_{ij} + \delta_{ijk} + t_{ijk.}/m, \text{ em que:}$$

$\bar{Y}_{ijk.}$: observação média (de m medições) do indivíduo ijk.

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$.

p_i : efeito da progénie ou família i, aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$.

b_j : efeito de bloco j, fixo, $E(b_j) = b_j$ e $E(b_j^2) = V_b$.

e_{ij} : efeito da parcela ij, aleatório, $E(e_{ij}) = 0$ e $E(e_{ij}^2) = \sigma_e^2$.

δ_{ijk} : efeito do indivíduo k dentro da parcela ij, aleatório, $E(\delta_{ijk}) = 0$ e $E(\delta_{ijk}^2) = \sigma_\delta^2$.

$t_{ijk.}/m$: efeito residual (ambiental) temporário médio (de m medições) associado ao indivíduo ijk, aleatório, $E(t_{ijk.}/m) = 0$ e $E(t_{ijk.}^2/m) = \sigma_t^2/m$.

No contexto do melhoramento genético de espécies perenes, os componentes de variância apresentados têm os seguintes significados (ignorando a epistasia):

σ_p^2 e σ_e^2 : variâncias entre progênies e entre parcelas, respectivamente.

$\sigma_\delta^2 = (1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2 + \sigma_{ep}^2$: variância permanente (de uma medição para outra) dentro de parcelas.

$\sigma_t^2 = \sigma_{et}^2$: variância ambiental temporária dentro de parcelas (variância entre medições dentro de indivíduo dentro de parcelas).

σ_a^2 e σ_d^2 : variâncias genética aditiva e de dominância, respectivamente.

σ_{ep}^2 : variância ambiental permanente.

$\sigma_{ad}^2 = \sigma_{ep}^2 + \sigma_{et}^2$: variância ambiental dentro de parcelas.

$\sigma_y^2 = \sigma_p^2 + \sigma_e^2 + \sigma_d^2 + \sigma_{et}^2$: variância fenotípica individual.

ρ_a : correlação genética intraclasses entre indivíduos da progénie ($\rho_a = 0,25$ para progênies de meios-irmãos e $\rho_a = 0,5$ para progênies de irmãos germanos).

ρ_d : coeficiente de parentesco de dominância entre os indivíduos da progénie ($\rho_d = 0$ para meios-irmãos e $\rho_d = 0,25$ para irmãos germanos).

7.4.1. Seleção por valores genéticos aditivos (propagação sexuada)

Nesta situação, o método de seleção índice multiefeitos (que considera todos os efeitos do modelo estatístico, exceto o efeito de bloco, o qual não afeta o mérito genético) para predição dos efeitos aditivos equivale a:

$I_1 = b_1(\bar{Y}_{ijK.} - \bar{Y}_{ij..}) + b_2(\bar{Y}_{i...} - \bar{Y}_{...}) + b_3(\bar{Y}_{ij.} - \bar{Y}_{i..} + \bar{Y}_{j..})$, em que:

$$b_1 = h_{dm}^2 = \frac{(1-\rho_a) \sigma_a^2}{(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2 + \sigma_{ep}^2 + \sigma_{et}^2/m} = \frac{(1-\rho_a) h_a^2}{[1+(m-1) \rho - mt_1]/m};$$

$$b_2 = h_{mm}^2 = \frac{\{[1+(nb-1) \rho_a]/nb\} \sigma_a^2}{\sigma_p^2 + \sigma_e^2/b + [(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2 + \sigma_{ep}^2]/(nb) + \sigma_{et}^2/(nbt)} =$$

$$= \frac{[1+(nb-1) \rho_a] h_a^2}{[1+(m-1) \rho + (n-1) mt_1 + (b-1)nmt_2]/m};$$

$$b_3 = h_{pm}^2 = \frac{[(1-\rho_a) \sigma_a^2]/n}{\sigma_e^2 + [(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2 + \sigma_{ep}^2]/n + \sigma_{et}^2/(nm)} =$$

$$= \frac{[1-\rho_d] h_a^2}{[1+(m-1) \rho + (n-1) mt_1 - nmt_2]/m}.$$

$h_a^2 = \frac{\sigma_a^2}{\sigma_y^2}$: herdabilidade no sentido restrito ao nível de indivíduos.

$\rho = \frac{\sigma_a^2 + \sigma_d^2 + \sigma_{ep}^2 + \sigma_e^2}{\sigma_y^2}$: repetibilidade individual ou correlação fenotípica intraclasse entre medições repetidas em um mesmo indivíduo.

$c^2 = \frac{\sigma_e^2}{\sigma_y^2}$: fração da correlação intraclasse entre os indivíduos da parcela atribuída ao efeito de ambiente comum aos indivíduos da parcela, ou seja, correlação entre indivíduos devida ao ambiente comum da parcela.

$t_1 = \frac{r_g \sigma_A^2 + \mu_g \sigma_D^2 + \sigma_e^2}{\sigma_F^2} = t_2 + c^2 =$ correlação fenotípica intraclasse entre os indivíduos da parcela.

$t_2 = \frac{\rho_a \sigma_a^2 + \rho_d \sigma_d^2}{\sigma_y^2} = \rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)$: correlação fenotípica intraclasse entre os indivíduos de uma progênie em diferentes blocos.

$h_g^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_y^2}$: herdabilidade no sentido amplo ao nível de indivíduos.

h_{dm}^2, h_{mm}^2 e h_{pm}^2 : herdabilidades dentro de parcela em nível de médias de indivíduos, entre médias de progênies e entre médias de parcelas, respectivamente.

$\bar{Y}_{ijK.}, \bar{Y}_{ij..}, \bar{Y}_{i...}, \bar{Y}_{j..}$ e $\bar{Y}_{...}$: médias do indivíduo, de parcela, de progênies, de bloco e geral, respectivamente.

Assim, para uma avaliação genética completa com progênie de meios-irmãos ou irmãos germanos são necessários os parâmetros: h_a^2 , ρ , h_g^2 e c^2 . Para progênies de meios-irmãos, quando o objetivo é a predição apenas dos efeitos aditivos, não é necessário o parâmetro h_g^2 . Para delineamentos com uma planta por parcela, tem-se $c^2 = 0$. É importante relatar que os modelos devem ser ajustados com o menor número possível de parâmetros. Uma vez que os procedimentos de predição assumem que as variâncias e covariâncias são conhecidas exatamente e na prática estes componentes são estimados, quanto menor o número de parâmetros utilizados menores erros estarão associados às soluções para os valores genéticos.

No contexto dos modelos lineares mistos, tem-se o modelo:

$$y = X\beta + Za + Wc + Tp + e, \text{ em que:}$$

y , β , a , c , p e e : vetores de dados ao nível de cada medição, de efeitos fixos (médias de blocos), de valores genéticos aditivos (aleatório), de efeitos de parcelas (aleatório), de efeitos permanentes (aleatório) e de erros aleatórios, respectivamente.

X , Z , W e T : matrizes de incidência para β , a , c e p , respectivamente.

As equações de modelo misto para predição dos valores genéticos pelo procedimento BLUP sob modelo individual (que devem fornecer os mesmos resultados que o índice I_1 em situação de dados balanceados) equivalem a:

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

$$\lambda_1 = \frac{\sigma_{et}^2}{\sigma_a^2} = \frac{(1-\rho)}{h_a^2}; \quad \lambda_2 = \frac{\sigma_{et}^2}{\sigma_e^2} = \frac{(1-\rho)}{c^2}; \quad \lambda_3 = \frac{\sigma_{et}^2}{\sigma_{ep}^2 + \sigma_D^2} = \frac{(1-\rho)}{(\rho - h_a^2 - c^2)}.$$

Estes componentes de variância podem ser obtidos a partir dos parâmetros h_a^2 , ρ , c^2 e σ_y^2 , da seguinte maneira:

$$\sigma_a^2 = h_a^2 \sigma_y^2; \quad \sigma_e^2 = c^2 \sigma_y^2; \quad (\sigma_{ep}^2 + \sigma_D^2) = (\rho - h_a^2 - c^2) \sigma_y^2; \quad \sigma_{et}^2 = (1-\rho) \sigma_y^2.$$

Pelo modelo e equações apresentadas, o vetor p de efeitos permanentes contempla os efeitos de ambiente permanente somados aos efeitos de dominância. Assim, este modelo considera que toda a variação de dominância encontra-se dentro de progênies, o que é incorreto para progênies de irmãos germanos.

7.4.2. Seleção por valores genotípicos (propagação assexuada)

A predição dos efeitos de dominância pode ser realizada empregando-se o índice multiefeitos por meio de:

$$I_2 = b_4(\bar{Y}_{ijK} - \bar{Y}_{ij..}) + b_5(\bar{Y}_{i...} - \bar{Y}_{....}) + b_6(\bar{Y}_{ij.} - \bar{Y}_{i...} - \bar{Y}_{j..} + \bar{Y}_{....}), \text{ em que:}$$

$$b_4 = \frac{(1-\rho_d) \sigma_d^2}{(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2 + \sigma_{ep}^2 + \sigma_{et}^2 / m} = \frac{(1-\rho_d) (h_g^2 - h_a^2)}{[1 + (m-1) \rho - mt_1] / m};$$

$$b_5 = \frac{\{ [1 + (nb-1) \rho_d] / nb \} \sigma_d^2}{\sigma_p^2 + \sigma_e^2 / b + [(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2 + \sigma_{ep}^2] / (nb) + \sigma_{et}^2 / (nbt)} =$$

$$= \frac{[1 + (nb-1) \mu_g] (h_g^2 - h_a^2)}{[1 + (m-1) \rho + (n-1) mt_1 + (b-1) nmt_2] / m};$$

$$b_6 = \frac{[(1-\rho_d) \sigma_d^2] / n}{\sigma_e^2 + [(1-\rho_a) \sigma_a^2 + (1-\rho_d) \sigma_d^2 + \sigma_{ep}^2] / n + \sigma_{et}^2 / (nm)} =$$

$$= \frac{[1-\rho_d] (h_g^2 - h_a^2)}{[1 + (m-1) \rho + (n-1) mt_1 - nmt_2] / m}.$$

No contexto dos modelos lineares mistos, tem-se o modelo:

$$y = X\beta + Za + Zd + Wc + Tp + e, \text{ em que } d \text{ é o vetor dos efeitos de dominância.}$$

As equações de modelo misto para a predição dos efeitos aditivos e de dominância pelo procedimento BLUP sob modelo individual equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{d} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'Z & X'W & X'T \\ ZX & ZZ + A^{-1}\lambda_1 & Z'Z & Z'W & ZT \\ ZX & Z'Z & Z'Z + D^{-1}\lambda_4 & Z'W & ZT \\ WX & W'Z & W'Z & WW + I\lambda_2 & WT \\ TX & TZ & TZ & TW & TT + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ Z'y \\ Z'y \\ T'y \end{bmatrix} \quad (2), \text{ em que:}$$

$$\lambda_4 = \frac{\sigma_{et}^2}{\sigma_d^2} = \frac{1-\rho}{(h_g^2 - h_a^2)}; \quad \lambda_5 = \frac{\sigma_{et}^2}{\sigma_{ep}^2} = \frac{1-\rho}{\rho - h_g^2 - c^2};$$

D : matriz de parentesco de dominância entre os indivíduos.

Os parâmetros σ_d^2 e σ_{ep}^2 podem ser obtidos por:

$$\sigma_d^2 = d^2 \quad \sigma_y^2 = (h_g^2 - h_a^2) \sigma_y^2$$

$\sigma_{ep}^2 = p^2 \sigma_y^2 = (\rho - h_g^2 - c^2) \sigma_y^2$, em que:

$d^2 = \frac{\sigma_d^2}{\sigma_y^2} = (h_g^2 - h_a^2)$: herdabilidade dos efeitos de dominância.

$p^2 = \frac{\sigma_{ep}^2}{\sigma_y^2} = \rho - h_g^2 - c^2$: correlação entre medidas repetidas devida ao ambiente permanente.

Os valores genotípicos preditos dos indivíduos são dados por $\hat{g} = \hat{a} + \hat{d}$ e sendo $Var(g) = A \sigma_a^2 + D \sigma_d^2 = G$, estes podem ser preditos diretamente pelas equações de modelo misto:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'T \\ Z'X & Z'Z + G^{-1}\sigma_{et}^2 & Z'W & Z'T \\ W'X & W'Z & W'W + I\lambda_2 & W'T \\ T'X & T'Z & T'W & T'T + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \end{bmatrix} \quad (3)$$

Os efeitos genotípicos preditos diretamente pelo índice multiefeitos são dados por:

$$\begin{aligned} I_3 &= b_7(\bar{Y}_{ijK} - \bar{Y}_{ij.}) + b_8(\bar{Y}_{i..} - \bar{Y}_{...}) + b_9(\bar{Y}_{ij..} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}) \\ &= (b_1 + b_4)(\bar{Y}_{ijK} - \bar{Y}_{ij..}) + (b_2 + b_5)(\bar{Y}_{i..} - \bar{Y}_{...}) + (b_3 + b_6)(\bar{Y}_{ij..} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}) \end{aligned}$$

O índice I_3 pode também ser aplicado à seleção de clones em testes clonais, bastando considerar $\rho_a = \rho_d = 1$.

7.4.3. Exemplo prático

Considerou-se, para exercício das metodologias, os seguintes componentes de variância:

$$\hat{\sigma}_p^2 = 0,2722 \quad \hat{\sigma}_e^2 = 0,8713 \quad \hat{\sigma}_\delta^2 = 1,9702 \quad \hat{\sigma}_{et}^2 = 4,3610 \quad \hat{\sigma}_y^2 = 7,4747 \quad \hat{\sigma}_a^2 = 1,0888 \cdot$$

$$\hat{\sigma}_d^2 = 0,5444 \quad \hat{\sigma}_{ep}^2 = 0,6092 \cdot$$

Estes componentes de variância estão associados aos seguintes parâmetros:

$$\hat{h}_a^2 = 0,1457; \quad \hat{h}_g^2 = 0,2185; \quad \hat{\rho} = 0,4166; \quad \hat{c}^2 = 0,1166.$$

$$\lambda_1 = 4,0053 \quad \lambda_2 = 5,0052 \quad \lambda_3 = 3,7803 \quad \lambda_4 = 8,01065 \quad \lambda_5 = 7,1586$$

Para aplicação das metodologias considerou-se o seguinte subconjunto de dados associados a um teste de progêniés de meios-irmãos ($\rho_a = 0,25$ e $\rho_d = 0$) de *Eucalyptus grandis*.

Observações	Indivíduos	Bloco	Progênie (Matriz)	Árvore	Medição	Diâmetro (cm)
1	1	1	1	1	1	9,87
2	1	1	1	1	2	10,00
3	2	1	1	2	1	14,48
4	2	1	1	2	2	15,00
5	3	1	2	1	1	8,91
6	3	1	2	1	2	9,00
7	4	1	2	2	1	15,00
8	4	1	2	2	2	14,64
9	5	2	1	1	1	9,55
10	5	2	1	1	2	10,00
11	6	2	1	2	1	8,00
12	6	2	1	2	2	7,96
13	7	2	2	1	1	15,00
14	7	2	2	1	2	16,07
15	8	2	2	2	1	14,01
16	8	2	2	2	2	14,50

Os vetores e matrizes apresentam as seguintes dimensões neste exemplo:

$y = (16 \times 1)$; $\beta = (2 \times 1)$; $a = d = (10 \times 1)$, incluindo as duas matrizes (genitoras). $c = (4 \times 1)$; $p = (8 \times 1)$; $X = (16 \times 2)$; $Z = (16 \times 10)$; $W = (16 \times 4)$; $T = (16 \times 8)$.

A matriz A (10×10) apresenta valores 1 na diagonal, 0,5 nas células correspondentes aos parentescos mães-filhos, 0,25 nas células correspondentes aos parentescos de meios-irmãos e 0 nas demais células.

A matriz de parentesco de dominância é a identidade, visto que os indivíduos e matrizes, neste caso, não possuem parentesco de dominância.

A seguir são apresentados os resultados da aplicação prática.

Seleção por valores genéticos aditivos

Os resultados referentes ao índice I_1 e método BLUP(1) são apresentados na Tabela 7.5.

Tabela 7.5. Coeficientes de ponderação (b_i) dos diferentes efeitos do modelo estatístico e valores genéticos preditos pelo índice I_1 e método BLUP(1).

b_1	Indivíduos	Valores genéticos preditos pelo índice I_1	Valores genéticos preditos pelo BLUP(1)
$\hat{b}_1 = 0,1967$	1	-0,6285	-0,6285
$\hat{b}_2 = 0,2729$	2	0,3169	0,3169
$\hat{b}_3 = 0,1386$	3	-0,4211	-0,4211

Continua...

Tabela 7.5. Continuação.

b_1	Indivíduos	Valores genéticos preditos pelo índice I_1	Valores genéticos preditos pelo BLUP(1)
	4	0,7327	0,7327
	5	-0,4273	-0,4273
	6	-0,7805	-0,7805
	7	0,7298	0,7298
	8	0,4780	0,4780
	9*	-	-0,4341
	10**	-	0,4341

* e ** - correspondem às matrizes 1 e 2, respectivamente.

Verifica-se a concordância entre os métodos I_1 e BLUP (1) para o caso balanceado. Adicionalmente, o BLUP (1) forneceu também as previsões dos valores genéticos das matrizes (Tabela 7.5).

Os demais efeitos estimados e preditos pelo BLUP (1) são apresentados na Tabela 7.6.

Tabela 7.6. Efeitos fixos de blocos ($\hat{\beta}_1$ e $\hat{\beta}_2$) e efeitos de parcela (\hat{c}) e permanentes (\hat{p}) estimados e preditos pelo método BLUP (1).

$\hat{\beta}$	\hat{c}	\hat{p}
$\hat{\beta}_1 = \bar{Y}_{1..} = 12,1125$	$\hat{c}_1 = 0,1307$	$\hat{p}_1 = -0,5812$
$\hat{\beta}_2 = \bar{Y}_{2..} = 11,8862$	$\hat{c}_2 = -0,1307$	$\hat{p}_2 = 0,7543$
	$\hat{c}_3 = -0,8255$	$\hat{p}_3 = -0,9016$
	$\hat{c}_4 = 0,8255$	$\hat{p}_4 = 0,7285$
		$\hat{p}_5 = -0,2970$
		$\hat{p}_6 = -0,7959$
		$\hat{p}_7 = 0,7243$
		$\hat{p}_8 = 0,3686$

Seleção por valores genotípicos

Os resultados referentes ao método BLUP (2) são apresentados na Tabela 7.7.

Tabela 7.7. Efeitos fixos e aleatórios estimados e preditos pelo procedimento BLUP (2).

Efeitos	Soluções
Efeitos fixos	
$\hat{\beta}_1$	12,1125
$\hat{\beta}_2$	11,8862
Efeitos aleatórios	
Efeitos aditivos	
\hat{a}_1	-0,6285
\hat{a}_2	0,3169
\hat{a}_3	-0,4211
\hat{a}_4	0,7327
\hat{a}_5	-0,4273
\hat{a}_6	-0,7805
\hat{a}_7	0,7298
\hat{a}_8	0,4780
\hat{a}_9	-0,4341
\hat{a}_{10}	0,4341
Efeitos de dominância	
\hat{d}_1	-0,2743
\hat{d}_2	0,3559
\hat{d}_3	-0,4255
\hat{d}_4	0,3438
\hat{d}_5	-0,1402
\hat{d}_6	-0,3756
\hat{d}_7	0,3418
\hat{d}_8	0,1739
\hat{d}_9	0,0000
\hat{d}_{10}	0,0000
Efeitos de ambiente permanente	
\hat{p}_1	-0,3069
\hat{p}_2	0,3983
\hat{p}_3	-0,4761

Continua...

Tabela 7.7. Continuação

Efeitos	Soluções
\hat{P}_4	0,3847
\hat{P}_5	-0,1569
\hat{P}_6	-0,4203
\hat{P}_7	0,3825
\hat{P}_8	0,1947
Efeitos de parcela	
\hat{c}_1	0,1307
\hat{c}_2	-0,1307
\hat{c}_3	-0,8255
\hat{c}_4	0,8255

Os resultados obtidos para \hat{a} e \hat{d} foram idênticos aos obtidos pelos índices I_1 e I_2 , respectivamente, conforme esperado. Os coeficientes de ponderação do Índice I_2 foram: $\hat{b}_4 = 0,1312$; $\hat{b}_5 = 0,0780$ e $\hat{b}_6 = 0,0923$.

Os valores de \hat{p} apresentados na Tabela 7.7 diferem daqueles apresentados na Tabela 7.6, cujos efeitos permanentes preditos equivalem à soma dos efeitos de ambiente permanente e efeitos de dominância, ao passo que na Tabela 7.7 referem-se apenas aos efeitos de ambiente permanente.

Na Tabela 7.8 são apresentados os valores genotípicos preditos pelos procedimentos I_3 e BLUP (3).

Tabela 7.8. Valores genotípicos ($\hat{g} = \hat{a} + \hat{d}$) preditos pelos procedimentos I_3 e BLUP(3).

Indivíduos	\hat{g}
1	-0,9027
2	0,6728
3	-0,8466
4	1,0765
5	-0,5675
6	-1,1561
7	1,0716
8	0,6519
9	-0,4341
10	0,4341

Os valores genotípicos preditos dos indivíduos (matrizes) 9 e 10 equivalem a seus valores genéticos aditivos preditos, pois estes não possuem observações próprias que permitam estimar os seus efeitos de dominância. Os valores genotípicos preditos pelos métodos I_3 e BLUP (3) são equivalentes e os ponderadores dos efeitos do modelo estatístico, pelo método I_3 , foram: $\hat{b}_7 = 0,3279$, $\hat{b}_8 = 0,3509$ e $\hat{b}_9 = 0,2309$.

Inclusão dos efeitos de medições

Para a situação de dados balanceados no que se refere às medições, a inclusão dos efeitos de medições não altera a estimação e predição dos efeitos de interesse. Incluindo os efeitos de medições como efeitos fixos no vetor β , verificou-se que a predição dos demais efeitos não foi alterada. O procedimento BLUP forneceu as seguintes estimativas dos efeitos de medições: -0,1469 para a medição 1 e 0,1469 para a medição 2. Assim, as médias das medições foram $11,9993 - 0,1469 = 11,8524$ para a medição 1 e $11,9993 + 0,1469 = 12,1462$ para a medição 2.

Entretanto, na situação de dados desbalanceados quanto ao número de medições e quando o efeito de medições for significativo, recomenda-se a inclusão deste efeito e do efeito da interação bloco x medição, no vetor β . Neste caso, os valores fenotípicos de indivíduos avaliados em uma só medição serão ajustados para a média do bloco na medição específica e não para a média global do bloco.

A inclusão dos efeitos de medições em β , em geral, conduz à singularidade da matriz dos coeficientes. Neste caso, recomenda-se a utilização do modelo de médias para os efeitos de bloco e a restrição somatório zero para os efeitos de medições para que $X'X$ e, consequentemente, a matriz dos coeficientes tenham inversa comum. Como $X\hat{\beta}$ é função estimável, ou seja, resulta em únicas estimativas BLUE (melhor estimação linear não viciada) dos efeitos fixos qualquer que seja o $\hat{\beta}$ utilizado (Graybill, 1976; Milliken & Johnson, 1984), estas restrições conduzem, essencialmente, aos mesmos resultados em $X\hat{\beta}$, os quais são utilizados para correção e ajuste dos valores fenotípicos individuais.

Na estimativa de componentes de variância por meio de métodos (REML, por exemplo) que iteragem nas equações de modelo misto, faz-se necessária a inclusão do efeito fixo de medição, mesmo em casos de dados balanceados quanto à medições. Caso contrário, a variância fenotípica será superestimada e a herdabilidade subestimada. Os efeitos fixos de blocos e medições, bem como o efeito fixo da interação bloco x medições podem ser ajustados em um único efeito denominado combinação bloco-medição, cujo vetor contém bm elementos. Neste caso, cada observação individual é ajustada para a média não viésada do bloco em uma dada medição e não haverá o problema de singularidade da matriz dos coeficientes. Dessa forma, recomenda-se o ajuste da combinação bloco-medição e não dos três efeitos mencionados.

Aplicação em espécies de crescimento contínuo

Na maioria das espécies florestais de rápido crescimento, caracteres tais quais a altura e o diâmetro são medidos anualmente a partir de um ano de idade. Entretanto, raramente, a seleção prática tem sido baseada em mais de uma avaliação. Os métodos descritos no presente trabalho permitem o uso de todas as medições realizadas visando à predição dos valores genéticos de cada árvore.

Neste caso, deve-se trabalhar com os incrementos anuais, empregando-se o modelo de repetibilidade. Os valores genéticos preditos (\hat{a}) dessa forma referem-se ao caráter incremento médio anual (para o caráter em questão), sendo que os valores genéticos para a própria altura ou diâmetro em uma determinada idade J , por exemplo, aos sete anos, são preditos por $\hat{a}_J = J\hat{a} = 7\hat{a}$ (desde que os incrementos anuais sejam aproximadamente de mesma magnitude). A média geral em tal idade pode ser dada pelo produto do incremento médio anual da população por J .

Uma grande vantagem desta metodologia refere-se à seleção de indivíduos avaliados em diferentes números de idades. Neste caso, por exemplo, indivíduos avaliados apenas ao 2º e 3º anos de idade serão penalizados em relação a indivíduos avaliados ao 2º, 3º e 4º anos de idade, em função do valor de m (número de medições) ser considerado no denominador dos coeficientes de ponderação b_1 a b_6 .

Uma abordagem mais completa para a estimativa e predição de caracteres de crescimento contínuo em várias idades pode ser encontrada no Capítulo 11.

Considerações finais

(i) As equações de modelo misto apresentadas permitem uma acurada predição de valores genéticos e genotípicos em situações de dados desbalanceados e com medidas repetidas, e são, prontamente, aplicáveis e de grande utilidade prática em programas de melhoramento genético de espécies perenes; (ii) para pleno uso das equações de modelo misto apresentadas são necessárias estimativas apenas dos parâmetros repetibilidade, herdabilidade no sentido amplo, herdabilidade no sentido restrito e correlação devida ao ambiente comum da parcela.

7.5. BLUP Individual e Índice Multiefeitos em Presença de Interação Genótipo x Ambiente

O uso de delineamentos experimentais repetidos em vários ambientes ou locais é prática comum no melhoramento genético de plantas (Vencovsky & Barriga, 1992; Ramalho et al., 1993). A avaliação genética e seleção nesta situação são bastante complexa, sobretudo, em decorrência do fenômeno da interação

genótipo x ambiente. A consideração deste fenômeno nos métodos de seleção conduz à necessidade de derivação de preditores de valores genéticos e genotípicos para cada indivíduo em cada ambiente, de forma que cada indivíduo possua um valor genético ou genotípico predito para cada ambiente em que for utilizado. A seguir são apresentados os desenvolvimentos realizados por Resende (1999b) e Resende et al. (1999).

Na avaliação de p progênies em b blocos com n plantas por parcela, considerando o delineamento de blocos ao acaso repetido em ℓ locais, tem-se o seguinte modelo estatístico:

$$Y_{ijkl} = \mu + p_i + \ell_j + p\ell_{ij} + b_{kj} + e_{ijk} + \delta_{ijkl}, \text{ em que:}$$

Y_{ijkl} : observação no indivíduo $ijkl$.

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$.

p_i : efeito da progénie ou família i , aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$.

ℓ_j : efeito fixo do local j , $E(\ell_j) = \ell_j$ e $E(\ell_j^2) = V_\ell$.

$p\ell_{ij}$: efeito aleatório da interação da progénie i com o local j , $E(p\ell_{ij}) = 0$ e $E(p\ell_{ij}^2) = \sigma_{p\ell}^2$.

b_{kj} : efeito fixo do bloco ou repetição k dentro do local j , $E(b_{kj}) = b_{kj}$ e $E(b_{kj}^2) = b_{kj}^2$.

e_{ijk} : efeito aleatório da parcela ik dentro do local j , $E(e_{ijk}) = 0$ e $E(e_{ijk}^2) = \sigma_e^2$.

δ_{ijkl} : efeito do indivíduo ℓ da progénie i no bloco k do local j , aleatório, $E(\delta_{ijkl}) = 0$ e $E(\delta_{ijkl}^2) = \sigma_\delta^2$.

Os efeitos e_{ijk} apresentam distribuição normal e são independentemente distribuídos e independentes dos demais efeitos aleatórios. As mesmas considerações são válidas para os efeitos δ_{ijkl} .

Os componentes de variância apresentados têm os seguintes significados (ignorando a epistasia):

σ_p^2 , $\sigma_{p\ell}^2$, σ_e^2 e σ_δ^2 : variâncias genética entre progênies, da interação progênies x locais, entre parcelas e dentro de parcelas, respectivamente.

$$\sigma_\delta^2 = (1 - \rho_a) \sigma_{aj}^2 + (1 - \rho_d) \sigma_{dj}^2 + \sigma_{adj}^2$$

σ_{aj}^2 e σ_{dj}^2 : variâncias genética aditiva e de dominância do local j , respectivamente.

σ_{adj}^2 : variância ambiental dentro de parcelas no local j .

$\sigma_{rdj}^2 = \sigma_\delta^2 - (1 - \rho_a) \sigma_{aj}^2$: variância residual dentro de parcelas no local j .

ρ_a : correlação genética intraclasse entre indivíduos da progénie ($\rho_a = 0,25$ para progênies de meios-irmãos e $\rho_a = 0,5$ para progênies de irmãos germanos).

ρ_d : coeficiente de parentesco de dominância entre os indivíduos da progênie ($\rho_d = 0$ para meios-irmãos e $\rho_d = 0,25$ para irmãos germanos).

Os componentes de variância associados ao modelo apresentado permitem a obtenção dos parâmetros para o local j :

$$\sigma_{yj}^2 = \sigma_{pj}^2 + \sigma_{ej}^2 + \sigma_{\delta_j}^2 = \text{variância fenotípica do local } j.$$

$$h_{aj}^2 = \frac{\sigma_{aj}^2}{\sigma_{yj}^2} : \text{herdabilidade no sentido restrito no local } j.$$

$$h_{gj}^2 = \frac{\sigma_{aj}^2 + \sigma_{dj}^2}{\sigma_{yj}^2} : \text{herdabilidade no sentido amplo no local } j.$$

$$c_j^2 = \frac{\sigma_{ej}^2}{\sigma_{yj}^2} : \text{correlação intraclasses devida ao ambiente comum da parcela no local } j.$$

Os componentes de variância envolvendo os locais, tomados dois a dois, permitem a obtenção dos parâmetros:

σ_a^2 e σ_d^2 : variâncias genética aditiva e de dominância, livres da interação genótipo x ambiente.

$r_a = \frac{\sigma_a^2}{\sigma_{a_1} \sigma_{a_2}}$: correlação genética aditiva entre o desempenho dos indivíduos de um local para outro (Falconer, 1952).

$r_d = \frac{\sigma_d^2}{\sigma_{d_1} \sigma_{d_2}}$: correlação genética de dominância entre o desempenho dos indivíduos de um local para outro.

7.5.1. Seleção por valores genéticos aditivos (propagação sexuada)

No caso do delineamento de blocos casualizados repetidos em vários locais, o método de seleção índice multiefeitos para a predição dos valores genéticos pode ser elaborado da forma relatada a seguir. Esses índices são apresentados considerando o local em que estão sendo avaliados os indivíduos e em que se pretende utilizá-los após a seleção.

Efeitos aditivos preditos para o local 1, para indivíduos avaliados no local 1

$I_1 = b_1(Y_{ilk} - \bar{Y}_{il.}) + b_2(\bar{Y}_{il..} - \bar{Y}_{..}) + b_3(\bar{Y}_{i2..} - \bar{Y}_{2..}) + b_4(\bar{Y}_{ilk} - \bar{Y}_{il..} - \bar{Y}_{lk..} + \bar{Y}_{..})$, em que:

$$b_1 = h_{\delta_1}^2 = \frac{(1-\rho_a) \sigma_{a_1}^2}{\sigma_{\delta_1}^2} = \frac{(1-\rho_a) h_{a1}^2}{1-t_{11}} \quad b_4 = h_{p_1}^2 = \frac{[(1-\rho_a)/n] \sigma_{a_1}^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2 / n} = \frac{[1-\rho_a] h_{a1}^2}{1+(n-1)t_{11}-nt_{21}}$$

b_2 e b_3 são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{11} \\ A_{12} \end{bmatrix}, \text{ em que:}$$

$$F_{11} = \sigma_{p_1}^2 + \sigma_{e_1}^2 / b + \sigma_{\delta_1}^2 / (nb) = \frac{1 + (n-1) t_{11} + (b-1) nt_{21}}{nb} \sigma_{y_1}^2$$

$$F_{12} = F_{21} = \rho_a \sigma_a^2 + \rho_d \sigma_d^2 = \rho_a r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2} + \rho_d r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}$$

$$F_{22} = \sigma_{p_2}^2 + \sigma_{e_2}^2 / r + \sigma_{\delta_2}^2 / (nb) = \frac{1 + (n-1) t_{12} + (b-1) nt_{22}}{nb} \sigma_{y_2}^2$$

$$A_{11} = \frac{1 + (nb-1) \rho_a}{nb} \sigma_{a_1}^2 = \frac{1 + (nb-1) \rho_a}{nb} h_{a1}^2 \sigma_{y_1}^2$$

$$A_{12} = \rho_a \sigma_a^2 = \rho_a r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}$$

$$t_{11} = \frac{\rho_a \sigma_{a_1}^2 + \rho_d \sigma_{d_1}^2 + \sigma_{e_1}^2}{\sigma_{y_1}^2} = t_{21} + c_1^2$$

$$t_{21} = \frac{\rho_a \sigma_{a_1}^2 + \rho_d \sigma_{d_1}^2}{\sigma_{y_1}^2} = \rho_a h_{a1}^2 + \rho_d (h_{g_1}^2 - h_{a1}^2)$$

$$t_{12} = \frac{\rho_a \sigma_{a_2}^2 + \rho_d \sigma_{d_2}^2 + \sigma_{e_2}^2}{\sigma_{y_2}^2} = t_{22} + c_2^2$$

$$t_{22} = \frac{\rho_a \sigma_{a_2}^2 + \rho_d \sigma_{d_2}^2}{\sigma_{y_2}^2} = \rho_a h_{a2}^2 + \rho_d (h_{g_2}^2 - h_{a2}^2)$$

$$D_1 = \left[\frac{\sigma_{d_1}^2}{\sigma_{y_1}^2} \right]^{1/2} = (h_{g_1}^2 - h_{a1}^2)^{1/2}$$

$$D_2 = \left[\frac{\sigma_{d_2}^2}{\sigma_{y_2}^2} \right]^{1/2} = (h_{g_2}^2 - h_{a2}^2)^{1/2}$$

$h_{\delta_1}^2$ e $h_{p_1}^2$: herdabilidades dentro de parcela e de parcela, respectivamente, para o local 1.

$\bar{Y}_{i1k\ell}$, $\bar{Y}_{i1k..}$, $\bar{Y}_{i1..}$, $\bar{Y}_{1k..}$ e $\bar{Y}_{1..}$: valor individual no local 1, média da parcela no local 1, média da progénie no local 1, média do bloco no local 1 e média geral no local 1, respectivamente.

$\bar{Y}_{i2..}$ e $\bar{Y}_{2..}$: médias da progénie e geral, respectivamente, no local 2.

Assim, para uma avaliação genética completa com progêneres de meios-irmãos ou irmãos germanos são necessários os parâmetros h_a^2 , h_g^2 e c^2 em cada local e r_a e r_d entre locais. Para progêneres de meios-irmãos, quando o objetivo é a predição apenas dos efeitos aditivos, não são necessários os parâmetros h_g^2 e r_d . Para delineamentos com uma planta por parcela tem-se $c^2 = 0$.

Efeitos aditivos preditos para o local 1, para indivíduos avaliados no local 2

$$I_2 = b_5(Y_{i2k\ell} - \bar{Y}_{i2k}) + b_6(\bar{Y}_{il..} - \bar{Y}_{l..}) + b_7(\bar{Y}_{i2..} - \bar{Y}_{2..}) + b_8(\bar{Y}_{i2k..} - \bar{Y}_{i2..} - \bar{Y}_{2k..} + \bar{Y}_{2..}), \text{ em que:}$$

$$b_5 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_{\delta_2}^2} = \frac{(1-\rho_a) r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{(1-t_{12}) \sigma_{y_2}^2}$$

$$b_8 = \frac{[(1-\rho_a)/n] \sigma_a^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2/n} = \frac{[(1-\rho_a)/n] r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{[1+(n-1)t_{12}-nt_{22}] \sigma_{y_2}^2}$$

b_6 e b_7 são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_6 \\ b_7 \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{11}^* \\ A_{12}^* \end{bmatrix}, \text{ em que:}$$

$$A_{11}^* = \rho_a \sigma_{a1}^2 = \rho_a h_{a1}^2 \sigma_{y_1}^2$$

$$A_{12}^* = \frac{1+(nb-1)}{nb} \rho_a \sigma_a^2 = \frac{1+(nb-1)}{nb} \rho_a r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}$$

Efeitos aditivos preditos para o local 2, para indivíduos avaliados no local 1

$$I_3 = b_9(Y_{ilk\ell} - \bar{Y}_{ilk}) + b_{10}(\bar{Y}_{il..} - \bar{Y}_{l..}) + b_{11}(\bar{Y}_{i2..} - \bar{Y}_{2..}) + b_{12}(\bar{Y}_{ilk..} - \bar{Y}_{il..} - \bar{Y}_{lk..} + \bar{Y}_{l..}), \text{ em que:}$$

$$b_9 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_{\delta_1}^2} = \frac{(1-\rho_a) r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{(1-t_{11}) \sigma_{y_1}^2}$$

$$b_{12} = \frac{[(1-\rho_a)/n] \sigma_a^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2/n} = \frac{[(1-\rho_a)/n] r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{[1+(n-1)t_{11}-nt_{21}] \sigma_{y_2}^2}$$

b_{10} e b_{11} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{10} \\ b_{11} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{21} \\ A_{22} \end{bmatrix}, \text{ em que:}$$

$A_{21} = A_{12}^*$, para o caso de sobrevivências iguais nos dois locais, caso contrário, *nb* refere-se ao número de plantas vivas para a progênie no local 1 para o cômputo de A_{21} , e ao número de plantas vivas para a progênie no local 2 para o cômputo de A_{12}^* .

$$A_{22} = \rho_a \sigma_{a_2}^2 = \rho_a h_{a_2}^2 \sigma_{y_2}^2$$

Efeitos aditivos preditos para o local 2, para indivíduos avaliados no local 2

$$I_4 = b_{13}(Y_{i2k\cdot} - \bar{Y}_{i2k\cdot}) + b_{14}(\bar{Y}_{i1\cdot} - \bar{Y}_{1\cdot}) + b_{15}(\bar{Y}_{i2\cdot} - \bar{Y}_{2\cdot}) + b_{16}(\bar{Y}_{i2k\cdot} - \bar{Y}_{i2\cdot} - \bar{Y}_{2k\cdot} + \bar{Y}_{2\cdot}), \text{ em que:}$$

$$b_{13} = h_{\delta_2}^2 = \frac{(1-\rho_a) \sigma_{a_2}^2}{\sigma_{\delta_2}^2} = \frac{(1-\rho_a) h_{a_2}^2}{1-t_{12}}$$

$$b_{16} = h_{p_2}^2 = \frac{[(1-\rho_a)/n] \sigma_{a_2}^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2/n} = \frac{(1-\rho_a) h_{a_2}^2}{1+(n-1)t_{12}-nt_{22}}$$

b_{14} e b_{15} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{14} \\ b_{15} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{21}^* \\ A_{22}^* \end{bmatrix}, \text{ em que:}$$

$$A_{21}^* = A_{12}$$

$$A_{22}^* = \frac{1+(nb-1)\rho_a}{nb} \sigma_{a_2}^2 = \frac{1+(nb-1)\rho_a}{nb} h_{a_2}^2 \sigma_{y_2}^2$$

No contexto dos modelos lineares mistos, tem-se o seguinte modelo:

$$y = X\beta + Za + Wc + e, \text{ em que:}$$

y , β , a , e c – vetores de dados, de efeitos fixos (médias de blocos nos 2 locais), de valores genéticos aditivos (aleatório) e de efeitos de parcelas (aleatório), respectivamente.

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

Considerando os dois locais, são os seguintes os modelos para cada local:

$$y_1 = X_1 \beta_1 + Z_1 a_1 + W_1 c_1 + e_1$$

$$y_2 = X_2 \beta_2 + Z_2 a_2 + W_2 c_2 + e_2$$

Adotando um modelo multivariado (Henderson & Quaas, 1976; Schaeffer et al., 1978), esses modelos podem ser expressos por:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Dessa forma, a_1 fornece os valores genéticos preditos de todos os indivíduos para o local 1 e a_2 apresenta os valores genéticos preditos de todos os indivíduos para o local 2.

Para o modelo bivariado, tem-se a seguinte estrutura de variâncias:

$$Var = \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \sigma_{a_1}^2 A & \sigma_a^2 A & 0 & 0 & 0 & 0 \\ \sigma_a^2 A & \sigma_{a_2}^2 A & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{e_1}^2 I & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{e_2}^2 I & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{rd_1}^2 I & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{rd_2}^2 I \end{bmatrix}$$

O modelo bivariado considera, então, o mesmo caráter avaliado em diferentes indivíduos nos diferentes locais, como sendo caracteres diferentes. Entretanto, em função da avaliação ser realizada em indivíduos diferentes, não existem covariâncias entre os efeitos de parcela nos dois locais e entre os efeitos residuais dentro de parcela nos dois locais.

As equações de modelo misto para predição dos valores genéticos pelo procedimento BLUP sob modelo individual (que devem fornecer simultaneamente os mesmos resultados dados pelo vetor formado pelos índices I_1 , I_2 , I_3 e I_4 , em situações de dados balanceados) equivalem a:

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

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}; \quad \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix}$$

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

$$R_O = \begin{bmatrix} \sigma_{rd_1}^2 & 0 \\ 0 & \sigma_{rd_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}$$

\otimes : denota a operação produto de Kronecker, no qual cada número da matriz multiplicadora (primeira matriz) deve multiplicar toda a segunda matriz.

Os componentes de variância podem ser obtidos a partir dos parâmetros h_a^2 , c^2 , σ_y^2 e r_a , da seguinte maneira:

$$\sigma_{aj}^2 = h_{aj}^2 \sigma_{yj}^2; \quad \sigma_{ej}^2 = c_j^2 \sigma_{yj}^2; \quad \sigma_{rdj}^2 = (1 - h_{aj}^2 - c_j^2) \sigma_{yj}^2; \quad \sigma_a^2 = r_a \sigma_{a_1} \sigma_{a_2}.$$

7.5.2. Seleção por valores genotípicos (propagação assexuada)

Para a predição dos efeitos de dominância, o método índice multiefeitos pode ser estruturado conforme relatado a seguir.

Efeitos de dominância preditos para o local 1, para indivíduos avaliados no local 1

$$I_5 = b_{17}(Y_{ilk} - \bar{Y}_{ilk}) + b_{18}(\bar{Y}_{il..} - \bar{Y}_{l..}) + b_{19}(\bar{Y}_{i2..} - \bar{Y}_{2..}) + b_{20}(\bar{Y}_{ilk} - \bar{Y}_{il..} - \bar{Y}_{lk..} + \bar{Y}_{l..}), \text{ em que:}$$

$$b_{17} = \frac{(1 - \rho_d) \sigma_{d_1}^2}{\sigma_{\delta_1}^2} = \frac{(1 - \rho_d) (h_{g_1}^2 - h_{al}^2)}{1 - t_{11}};$$

$$b_{20} = \frac{[(1 - \rho_d)/n] \sigma_{d_1}^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2 / n} = \frac{(1 - \rho_d) (h_{g_1}^2 - h_{al}^2)}{1 + (n-1) t_{11} - nt_{21}}$$

$$\begin{bmatrix} b_{18} \\ b_{19} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{11} \\ D_{12} \end{bmatrix}, \text{ em que:}$$

$$D_{11} = \frac{1 + (nb - 1) \rho_d}{nb} \sigma_{d_1}^2 = \frac{1 + (nb - 1) \rho_d}{nb} (h_{g_1}^2 - h_{al}^2) \sigma_{y_1}^2$$

$$D_{12} = \rho_d \sigma_d^2 = \rho_d r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}$$

Efeitos de dominância preditos para o local 1, para indivíduos avaliados no local 2

$$I_6 = b_{21}(Y_{i2kl} - \bar{Y}_{i2k..}) + b_{22}(\bar{Y}_{il..} - \bar{Y}_{l..}) + b_{23}(\bar{Y}_{i2..} - \bar{Y}_{2..}) + b_{24}(\bar{Y}_{i2k..} - \bar{Y}_{i2..} - \bar{Y}_{2k..} + \bar{Y}_{2..}), \text{ em que:}$$

$$b_{21} = \frac{(1 - \rho_d) \sigma_d^2}{\sigma_{\delta_2}^2} = \frac{(1 - \rho_d) r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{(1 - t_{12}) \sigma_{y_2}^2}$$

$$b_{24} = \frac{[(1 - \rho_d)/n] \sigma_d^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2 / n} = \frac{[(1 - \rho_d)/n] r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{[1 + (n-1) t_{12} - nt_{22}] \sigma_{y_2}^2}$$

b_{22} e b_{23} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{22} \\ b_{23} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{11}^* \\ D_{12}^* \end{bmatrix}, \text{ em que:}$$

$$D_{11}^* = \rho_d \sigma_{d_1}^2 = \rho_d (h_{g_1}^2 - h_{a1}^2) \sigma_{y_1}^2$$

$$D_{12}^* = \frac{1+(nb-1)}{nb} \rho_d \sigma_d^2 = \frac{1+(nb-1)}{nb} r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}$$

Efeitos de dominância preditos para o local 2, para indivíduos avaliados no local 1

$$I_7 = b_{25}(Y_{ilk\ell} - \bar{Y}_{ilk.}) + b_{26}(\bar{Y}_{il..} - \bar{Y}_{.1..}) + b_{27}(\bar{Y}_{i2..} - \bar{Y}_{.2..}) + b_{28}(\bar{Y}_{ilk.} - \bar{Y}_{il..} - \bar{Y}_{.lk.} + \bar{Y}_{.1..}), \text{ em que:}$$

$$b_{25} = \frac{(1-\rho_d) \sigma_d^2}{\sigma_{\delta_1}^2} = \frac{(1-\rho_d) r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{(1-t_{11}) \sigma_{y_1}^2}$$

$$b_{28} = \frac{[(1-\rho_d)/n] \sigma_d^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2 / n} = \frac{[(1-\rho_d)/n] r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{[1+(n-1) t_{11} - nt_{21}] \sigma_{y_2}^2}$$

b_{26} e b_{27} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{26} \\ b_{27} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{21} \\ D_{22} \end{bmatrix}, \text{ em que:}$$

$D_{21} = D_{12}^*$, para o caso de sobrevivências iguais nos dois locais.

$$D_{22} = \rho_d \sigma_{d_2}^2 = \rho_d (h_{g_2}^2 - h_{a2}^2) \sigma_{y_2}^2$$

Efeitos de dominância preditos para o local 2, para indivíduos avaliados no local 2

$$I_8 = b_{29}(Y_{i2k\ell} - \bar{Y}_{i2k.}) + b_{30}(\bar{Y}_{il..} - \bar{Y}_{.1..}) + b_{31}(\bar{Y}_{i2..} - \bar{Y}_{.2..}) + b_{32}(\bar{Y}_{i2k.} - \bar{Y}_{il..} - \bar{Y}_{.2k.} + \bar{Y}_{.2..}), \text{ em que:}$$

$$b_{29} = \frac{(1-\rho_d) \sigma_{d_2}^2}{\sigma_{\delta_2}^2} = \frac{(1-\rho_d) (h_{g_2}^2 - h_{a2}^2)}{1-t_{12}}$$

$$b_{32} = \frac{[(1-\rho_d)/n] \sigma_{d_2}^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2 / n} = \frac{(1-\rho_d) (h_{g_2}^2 - h_{a2}^2)}{1+(n-1) t_{12} - nt_{22}}$$

b_{30} e b_{31} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{30} \\ b_{31} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{21}^* \\ D_{22}^* \end{bmatrix}, \text{ em que:}$$

$$D_{21}^* = D_{12}$$

$$D_{22}^* = \frac{1+(nb-1)}{nb} \rho_d \quad \sigma_{d_2}^2 = \frac{1+(nb-1)}{nb} \rho_d (h_{g_2}^2 - h_{a2}^2) \sigma_{y_2}^2$$

No contexto dos modelos lineares mistos, adotando-se o modelo multivariado, tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \text{ em que } d_1$$

e d_2 referem-se aos vetores dos efeitos de dominância no local 1 e local 2, respectivamente.

As equações de modelo misto para a predição dos efeitos aditivos e de dominância pelo procedimento BLUP sob modelo individual (que devem fornecer os mesmos resultados dados pelo vetor formado pelos índices $I_1, I_2, I_3, I_4, I_5, I_6, I_7$ e I_8 , em situações de dados balanceados) equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{d} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'R_R^{-1}X & X'R_R^{-1}Z & X'R_R^{-1}Z & X'R_R^{-1}W \\ Z'R_R^{-1}X & Z'R_R^{-1}Z+G^{-1} & Z'R_R^{-1}Z & Z'R_R^{-1}W \\ Z'R_R^{-1}X & Z'R_R^{-1}Z & Z'R_R^{-1}Z+G_d^{-1} & Z'R_R^{-1}W \\ W'R_R^{-1}X & W'R_R^{-1}Z & W'R_R^{-1}Z & W'R_R^{-1}W+C^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R_R^{-1}y \\ Z'R_R^{-1}y \\ Z'R_R^{-1}y \\ W'R_R^{-1}y \end{bmatrix} \quad (2), \text{ em que:}$$

$$\hat{d} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \end{bmatrix}$$

$$R_R = R_{Ro}^{-1} \otimes I; \quad G_d^{-1} = G_{do}^{-1} \otimes D^{-1}$$

D : matriz de parentesco de dominância entre os indivíduos em avaliação.

$$R_{Ro} = \begin{bmatrix} \sigma_{ad_1}^2 & 0 \\ 0 & \sigma_{ad_2}^2 \end{bmatrix}; \quad G_{do} = \begin{bmatrix} \sigma_{D_1}^2 & \sigma_D^2 \\ \sigma_D^2 & \sigma_{D_2}^2 \end{bmatrix}$$

$$\sigma_{adj}^2 = \sigma_{\delta j}^2 - (1 - \rho_a) \sigma_{aj}^2 - (1 - \rho_d) \sigma_{dj}^2$$

Os componentes de variância podem ser obtidos a partir dos parâmetros h_a^2 , h_g^2 , c^2 , e r_d por meio de:

$$\sigma_{dj}^2 = (h_{gj}^2 - h_{aj}^2) \sigma_{Fj}^2; \quad \sigma_d^2 = r_d \sigma_{d_1} \sigma_{d_2}; \quad \sigma_{adj}^2 = (1 - h_{gj}^2 - c_j^2) \sigma_{yj}^2.$$

Os valores genotípicos preditos dos indivíduos são dados por: $\hat{g} = \hat{a} + \hat{d}$ e sendo $Var(g) = G_o \otimes A + G_{do} \otimes D = G_{ad}$, os valores genotípicos podem ser preditos diretamente por:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X' R_R^{-1} X & X' R_R^{-1} Z & X' R_R^{-1} W \\ Z' R_R^{-1} X & Z' R_R^{-1} Z + G_{ad}^{-1} & Z' R_R^{-1} W \\ W' R_R^{-1} X & W' R_R^{-1} Z & W' R_R^{-1} W + C^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X' R_R^{-1} y \\ Z' R_R^{-1} y \\ W' R_R^{-1} y \end{bmatrix} \quad (3)$$

Os valores genotípicos preditos diretamente pelo índice multiefeitos são obtidos conforme relatado a seguir:

(a) Valores genotípicos preditos para o local 1, para indivíduos avaliados no local 1.

$$I_9 = I_1 + I_5$$

(b) Valores genotípicos preditos para o local 1, para indivíduos avaliados no local 2.

$$I_{10} = I_2 + I_6$$

(c) Valores genotípicos preditos para o local 2, para indivíduos avaliados no local 1.

$$I_{11} = I_3 + I_7$$

(d) Valores genotípicos preditos para o local 2, para indivíduos avaliados no local 2.

$$I_{12} = I_4 + I_8$$

Os índices I_9 , I_{10} , I_{11} e I_{12} podem ser aplicados à seleção de clones em testes clonais, bastando considerar $\rho_a = \rho_d = 1$.

7.5.3. Exemplo prático

Para exercício das metodologias, foram considerados os seguintes componentes de variância:

Local 1:

$$h_{a1}^2 = 0,1457; \quad h_{g1}^2 = 0,2185; \quad c_1^2 = 0,1166; \quad \sigma_{y_1}^2 = 7,4747.$$

$$\begin{aligned} \sigma_{p_1}^2 &= 0,2722; \quad \sigma_{e_1}^2 = 0,8716; \quad \sigma_{\delta_1}^2 = 6,3307; \quad \sigma_{rd_1}^2 = 5,5141; \quad \sigma_{ad_1}^2 = 4,9699, \\ \sigma_{a_1}^2 &= 1,0888; \quad \sigma_{d_1}^2 = 0,5444. \end{aligned}$$

Local 2:

$$h_{a2}^2 = 0,1875; \quad h_{g2}^2 = 0,28125; \quad c_2^2 = 0,13; \quad \sigma_{y_2}^2 = 8,0.$$

$$\begin{aligned} \sigma_{p_2}^2 &= 0,375; \quad \sigma_{e_2}^2 = 1,04; \quad \sigma_{\delta_2}^2 = 6,585; \quad \sigma_{nd_2}^2 = 5,46; \quad \sigma_{ad_2}^2 = 4,71; \quad \sigma_{a_2}^2 = 1,5; \quad \sigma_{d_2}^2 = 0,75. \end{aligned}$$

Análise conjunta dos locais 1 e 2:

$$r_a = 0,7; \quad \sigma_a^2 = 0,8946$$

$$r_d = 0,8; \quad \sigma_d^2 = 0,5112$$

Para aplicação das metodologias, considerou-se o seguinte subconjunto de dados associados a um teste de progêniés de meios-irmãos ($\rho_a = 0,25$ e $\rho_d = 0$) de *Eucalyptus grandis*.

Indivíduo	Local	Bloco	Progênie	Árvore	Diâmetro (cm)
1	1	1	1	1	9,87
2	1	1	1	2	10,00
3	1	1	2	1	14,48
4	1	1	2	2	15,00
5	1	2	1	1	8,91
6	1	2	1	2	9,00
7	1	2	2	1	15,00
8	1	2	2	2	14,64
9	2	3	1	1	9,55
10	2	3	1	2	10,00
11	2	3	2	1	8,00
12	2	3	2	2	7,96
13	2	4	1	1	15,00
14	2	4	1	2	16,07
15	2	4	2	1	14,01
16	2	4	2	2	14,50

Os vetores e matrizes apresentam as seguintes dimensões neste exemplo:

$y = (32 \times 1)$; $f = (4 \times 1)$; $a = (36 \times 1)$ incluindo as duas matrizes (genitores); $d = (36 \times 1)$ incluindo as duas matrizes (genitores); $c = (8 \times 1)$; $X = (32 \times 4)$; $Z = (32 \times 36)$; $W = (32 \times 8)$; $A = (18 \times 18)$; $D = I_{(18)}$; $R^{-1} = R_{Ro}^{-1} \otimes I_{(16)}$; $C^{-1} = C_o^{-1} \otimes I_{(4)}$; $R_R^{-1} = R_{Ro}^{-1} \otimes I_{(16)}$.

É importante ressaltar que o vetor y apresenta seqüencialmente: 8 valores de diâmetro dos indivíduos do local 1, 16 valores zero e 8 valores de diâmetro dos indivíduos do local 2.

Na seqüência, são apresentados os resultados.

Seleção por valores genéticos aditivos

Os resultados referentes ao método BLUP (1), os quais são idênticos aos obtidos pelos índices I_1 a I_4 , são apresentados na Tabela 7.9.

Tabela 7.9. Coeficientes de ponderação (b_i) dos diferentes efeitos do modelo estatístico pelos métodos I_1 a I_4 e valores genéticos aditivos preditos por estes índices e pelo método BLUP (1).

b_i	Indivíduos	Efeitos Aditivos preditos para o local 1	Efeitos Aditivos preditos para o local 2
$b_1 = 0,1290$	1	-0,4639	-0,3031
$b_2 = 0,2011$	2	-0,4472	-0,2893
$b_3 = 0,0703$	3	0,4220	0,2686
$b_4 = 0,10114$	4	0,4891	0,3237
$b_5 = 0,10189$	5	-0,5149	-0,3450
$b_6 = 0,1047$	6	-0,5033	-0,3354
$b_7 = 0,1448$	7	0,5324	0,3593
$b_8 = 0,07743$	8	0,4859	0,3211
$b_9 = 0,10598$	9	-0,1809	-0,0230
$b_{10} = 0,1578$	10	-0,1350	0,0539
$b_{11} = 0,13368$	11	0,1600	-0,0120
$b_{12} = 0,0831$	12	0,1559	-0,0189
$b_{13} = 0,17084$	13	-0,2324	-0,1094
$b_{14} = 0,07305$	14	-0,1234	0,0734
$b_{15} = 0,25181$	15	0,1530	-0,0239
$b_{16} = 0,12983$	16	0,2029	0,0598
	17*	-0,4729	-0,2323
	18**	0,4729	0,2323

* e ** - correspondem às matrizes 1 e 2, respectivamente.

Verifica-se que um mesmo indivíduo apresenta dois valores genéticos preditos, um para cada local em que for utilizado (Tabela 7.9). Isto permite maior eficiência dos programas de melhoramento, em presença de interação genótipo x ambiente. Em função desta interação, cada indivíduo apresenta maior (em valor absoluto) efeito predito no local em que foi plantado experimentalmente.

Seleção por valores genotípicos

Os resultados referentes aos efeitos de dominância preditos pelo método BLUP (2), e índices I_5 a I_8 , são apresentados na Tabela 7.10.

Tabela 7.10. Coeficientes de ponderação (b_i) dos diferentes efeitos do modelo estatístico pelos métodos I_5 a I_8 e efeitos de dominância preditos por estes índices e pelo método BLUP (2).

b_1	Indivíduos	Efeitos de dominância preditos para o local 1	Efeitos de dominância preditos para o local 2
$b_{17} = 0,08599$	1	-0,1516	-0,1424
$b_{18} = 0,0599$	2	-0,1405	-0,1319
$b_{19} = -0,0053$	3	0,1237	0,1161
$b_{20} = 0,06743$	4	0,1684	0,1581
$b_{21} = 0,07763$	5	-0,1857	-0,1743
$b_{22} = -0,00495$	6	-0,1779	-0,1671
$b_{23} = 0,05073$	7	0,1973	0,1852
$b_{24} = 0,059$	8	0,1663	0,1562
$b_{25} = 0,08075$	9	0,0423	0,0621
$b_{26} = 0,0563$	10	0,0773	0,1134
$b_{27} = 0,0050$	11	-0,0582	-0,0855
$b_{28} = 0,0633$	12	-0,0614	-0,0900
$b_{29} = 0,1139$	13	0,0031	0,0045
$b_{30} = -0,0073$	14	0,0861	0,1264
$b_{31} = 0,0744$	15	-0,0636	-0,0934
$b_{32} = 0,08656$	16	-0,0256	-0,0375
	17*	0,0000	0,0000
	18**	0,0000	0,0000

* e ** - correspondem às matrizes 1 e 2, respectivamente.

Os demais efeitos estimados e preditos pelo BLUP (2), exceto os efeitos aditivos (que foram apresentados na Tabela 7.5.1), são apresentados na Tabela 7.11.

Tabela 7.11. Efeitos fixos de blocos ($\hat{\beta}_1$ a $\hat{\beta}_4$) e efeitos aleatórios de parcela (\hat{c}) e permanentes (\hat{p}) estimados e preditos pelo método BLUP(1).

$\hat{\beta}$	\hat{c}
$\hat{\beta}_1 = 12,3375$	$\hat{c}_1 = -0,4676$
$\hat{\beta}_2 = 11,8875$	$\hat{c}_2 = 0,4676$
$\hat{\beta}_3 = 8,8775$	$\hat{c}_3 = -0,5821$
$\hat{\beta}_4 = 14,8950$	$\hat{c}_4 = 0,5821$
	$\hat{c}_5 = 0,2433$
	$\hat{c}_6 = -0,2433$
	$\hat{c}_7 = 0,1815$
	$\hat{c}_8 = -0,1815$

Sendo os efeitos de blocos hierárquicos em relação ao efeito de locais, não há necessidade de incluir os efeitos de locais em β , já que β , neste caso, fornece a estimativa para todos os blocos ao longo dos locais.

Na Tabela 7.12 são apresentados os valores genotípicos preditos pelos métodos I_9 a I_{12} e BLUP(3).

Tabela 7.12. Valores genotípicos ($\hat{g} = \hat{a} + \hat{d}$) preditos pelos procedimentos I_9 a I_{12} e BLUP (3).

Indivíduos	\hat{g} no local 1	\hat{g} no local 2
1	-0,6156	-0,4454
2	-0,5876	-0,4212
3	0,5457	0,3848
4	0,6575	0,4819
5	-0,7006	-0,5193
6	-0,6812	-0,5025
7	0,7296	0,5445
8	0,6522	0,4773
9	-0,1386	0,0391
10	-0,0578	0,1672
11	0,1018	-0,0975
12	0,0946	-0,1089
13	-0,2293	-0,1049
14	-0,0373	0,1998
15	0,0893	-0,1172
16	0,1773	0,0223
17	-0,4729	-0,2323
18	0,4729	0,2323

Os resultados obtidos revelaram grande eficácia e flexibilidade do procedimento BLUP individual sob modelo multivariado para a predição de valores genéticos e genotípicos em presença de interação genótipo x ambiente.

7.5.4. Heterogeneidade de variâncias

O procedimento BLUP considera que os componentes de variância (matrizes G e R) são conhecidos, e que as propriedades dos preditores somente são asseguradas nestas condições (Henderson, 1984). Assim, na prática, os componentes de variância devem ser estimados da maneira mais fidedigna possível para que as estimativas possam substituir adequadamente os valores paramétricos. Neste caso, o procedimento padrão para estimação dos componentes de variância no contexto dos modelos lineares mistos é o da máxima verossimilhança restrita (REML) descrito em detalhes por Searle et al. (1992) e Rao (1997). Dependendo da forma como os componentes de variância são estima-

dos, existe uma semelhança formal entre o BLUP e a estimativa bayesiana empírica (Harville, 1977) e este procedimento de estimativa de componentes de variância/predição dos efeitos aleatórios, simultaneamente, é denominado BLUP empírico (Harville & Carriquiry, 1992).

Uma questão relevante associada à metodologia de modelos mistos refere-se a variâncias desiguais ou heterogêneas através dos níveis dos efeitos fixos. Estas variâncias heterogêneas estão associadas ao componente de variância da interação genótipo x ambiente (σ_{pe}^2) ou genótipo x níveis dos efeitos fixos. A variância da interação genótipo x ambiente pode ser desdobrada nas partes simples e complexa (Robertson, 1959), sendo que a parte simples é devida à heterogeneidade de variâncias nos ambientes e a parte complexa é devida à correlação genética menor que 1 através dos ambientes. Assim, quando a correlação genética através dos ambientes equivale a 1, a variância da interação genótipo x ambiente refere-se, exclusivamente, à heterogeneidade de variâncias. Neste caso, existem as seguintes possibilidades (Henderson, 1984; Garrick & Van Vleck, 1987; Visscher & Hill, 1992):

- (i) Homogeneidade de variância genética e heterogeneidade de variância ambiental e, portanto, herdabilidades diferentes nos ambientes.
- (ii) Heterogeneidade de variância genética e homogeneidade de variância ambiental e, portanto, herdabilidades diferentes nos ambientes.
- (iii) Variâncias genética e ambiental heterogêneas, mas mudando proporcionalmente nos ambientes de forma que a herdabilidade se mantém constante.
- (iv) Variâncias genética e ambiental heterogêneas conduzindo a herdabilidades variáveis nos ambientes.

Quando a correlação genética (r_a) através dos ambientes equivale a 1 e, tanto as variâncias genéticas quanto as variâncias ambientais são homogêneas (herdabilidades iguais nos ambientes e variância da interação igual a zero), as equações de modelo misto para o caso univariado são completamente adequadas. Na presença de heterogeneidade de variâncias, as equações de modelo misto (EMM) para o BLUP podem também ser utilizadas adequadamente (Henderson, 1984). Na situação (i), a matriz R dos resíduos não pode ser estruturada como $R = I\sigma_e^2$ e, portanto, as EMM para o caso univariado não podem ser fatoradas em função de R. Para as situações (ii), (iii), (iv) e quando a correlação genética através dos ambientes é menor que 1, torna-se necessário o uso do modelo multivariado conforme apresentado no presente trabalho. Assim, o modelo multivariado é recomendado não apenas quando $r_a < 1$, mas, também, quando as herdabilidades e/ou as variâncias genéticas diferem através dos ambientes. O BLUP multivariado considera adequadamente a questão da interação genótipo x ambiente e heterogeneidade de variâncias, permitindo, também, explorar as diferentes herdabilidades nos ambientes (as herdabilidades podem ser maiores em ambientes que permitem maior expressão da variabilidade genética e, neste caso, este fato deve ser capitalizado na seleção).

Embora o modelo multivariado seja o modelo correto para lidar com a heterogeneidade de variâncias e interação genótipo x ambiente, uma possível heterogeneidade de variância entre blocos dentro de locais não é levada em consideração. Este fato pode conduzir à seleção de maior número de indivíduos nos blocos mais variáveis fenotípicamente, o que é incorreto quando a herdabilidade nestes blocos não é superior. Assim, torna-se recomendável a transformação dos dados usando uma relação entre as variâncias fenotípicas populacionais (σ_p^2) e dentro de blocos (σ_B^2), conforme Visscher et al. (1991). Uma transformação recomendável é:

$$Y_{ij}^* = Y_i + [\sigma_p(Y_{ij} - Y_i) / s_{B_i}], \text{ em que:}$$

Y_{ij}^* : valor individual corrigido.

Y_{ij} : valor individual não corrigido.

Y_i : média do bloco i .

σ_p : desvio padrão populacional ou do local, que pode ser tomado como a média dos desvios padrões dos blocos.

s_{B_i} : desvio padrão fenotípico dentro do bloco i .

Se as precisões dos desvios padrões estimados para cada bloco diferem, recomenda-se (Visscher & Hill, 1992) que a obtenção de s_{B_i} seja baseada na regressão para uma estimativa geral a priori, que pode ser o próprio σ_p . Weigel & Gianola (1993) apresentaram um procedimento bayesiano empírico para cômputo de s_{B_i} , dado por:

$$\hat{s}_{B_i}^2 = \sigma_p^2 + \frac{d}{d+d_0} (\hat{s}_{B_i}^2 - \sigma_p^2) \text{ em que } d \text{ refere-se aos graus de liberdade da estimativa inicial } \hat{s}_{B_i}^2 \text{ da variância dentro de blocos e } d_0 \text{ aos graus de liberdade da estimativa } \sigma_p^2.$$

Quanto aos efeitos da padronização, Ibanez et al. (1996) mostraram que os valores genéticos preditos dos indivíduos avaliados em blocos com alta/baixa variância tendem a diminuir/aumentar quando a transformação para correção de heterogeneidade de variância é realizada.

Em geral, é sempre recomendável a padronização quando lidar com dados de vários experimentos ou de diferentes idades ou safras, a menos que se utilize o modelo multivariado. A divisão de cada observação pelo desvio padrão fenotípico do respectivo bloco ou da idade em questão também é uma forma eficiente de padronização e correção para a heterogeneidade de variância.

Considerações finais

(i) As equações de modelo misto apresentadas, associadas ao BLUP individual multivariado, permitem a predição de valores genéticos e genotípicos para cada indivíduo em cada local em que for utilizado, considerando de maneira eficiente a ocorrência de interação genótipo x ambiente; (ii) para pleno uso destas equações de modelo misto são necessárias estimativas dos parâmetros herdabilidade no sentido restrito, herdabilidade no sentido amplo e correlação devida ao ambiente comum da parcela em cada local e correlação genética aditiva e de dominância através dos locais.

7.6. BLUP Individual e Índice Multiefeitos em Presença de Interação Genótipo x Ambiente e Medidas Repetidas

Uma vez que os programas de melhoramento genético de espécies perenes visam gerar material genético melhorado para condições ambientais diversas (Vencovsky & Barriga, 1992; Ramalho et al., 1993), torna-se relevante a predição dos efeitos aditivos e de dominância dos indivíduos para os diferentes ambientes, considerando a ocorrência de interação genótipo x ambiente. Outros fatores a serem considerados nas espécies perenes são a prática de medições repetidas em cada indivíduo (tornando-se necessária a adoção de um modelo de repetibilidade), e a utilização da propagação assexuada dos indivíduos superiores. Assim sendo, para uma seleção eficiente, é preciso utilizar modelos multivariados com medidas repetidas, inclusão dos efeitos de dominância e consideração da interação genótipo x ambiente. Esses modelos usam, simultaneamente, os principais desenvolvimentos (Comstock, 1996) conceituais e práticos experimentados pela genética quantitativa e estatística. A seguir são apresentados os desenvolvimentos realizados por Resende (1999b).

Na avaliação de p progênies em b blocos com n plantas por parcela e m medições por planta, considerando o delineamento de blocos ao acaso repetido em ℓ locais, tem-se o seguinte modelo estatístico:

$$\bar{Y}_{ijkl} = \mu + p_i + \ell_j + p\ell_{ij} + b_{kj} + e_{ijk} + \delta_{ijkl} + t_{ijkl} / m, \text{ em que:}$$

\bar{Y}_{ijkl} : observação média (de m medições) no indivíduo $ijkl$.

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$.

p_i : efeito da progénie ou família i , aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$.

ℓ_j : efeito fixo do local j , $E(\ell_j) = \ell_j$ e $E(\ell_j^2) = v_\ell$.

$p\ell_{ij}$: efeito aleatório da interação da progénie i com o local j ,

$$E(p\ell_{ij}) = 0 \text{ e } E(p\ell_{ij}^2) = \sigma_{p\ell}^2.$$

b_{kj} : efeito fixo do bloco ou repetição k dentro do local j , $E(b_{kj}) = b_{kj}$ e $E(b_{kj}^2) = v_b$.

e_{ijk} : efeito aleatório da parcela ik dentro do local j , $E(e_{ijk}) = 0$ e $E(e_{ijk}^2) = \sigma_e^2$.

$\delta_{ijk\ell}$: efeito do indivíduo ℓ da progênie i no bloco k do local j , aleatório, $E(\delta_{ijk\ell}) = 0$ e $E(\delta_{ijk\ell}^2) = \sigma_\delta^2$.

t_{ijkl}/m : efeito residual (ambiental) temporário médio (de m medições) associado ao indivíduo $ijkl$, aleatório, $E(t_{ijkl}/m) = 0$ e $E(t_{ijkl}^2/m) = \sigma_t^2/m$.

Os componentes de variância apresentados têm os seguintes significados (ignorando a epistasia):

σ_p^2 , σ_{pl}^2 , σ_e^2 e σ_d^2 : variâncias genética entre progênies, da interação progênies x locais, entre parcelas e dentro de parcelas, respectivamente.

$\sigma_t^2 = \sigma_{et}^2$: variância ambiental temporária dentro da parcela (variância entre medições dentro de indivíduo dentro de parcela).

$\sigma_{\tilde{g}}^2 = (1 - \rho_a) \sigma_{aj}^2 + (1 - \rho_d) \sigma_{dj}^2 + \sigma_{epj}^2$: variância permanente (de uma medição para outra) dentro de parcelas, no local j .

σ_{aj}^2 e σ_{dj}^2 : variâncias genética aditiva e de dominância do local j , respectivamente.

σ_{epj}^2 : variância ambiental permanente dentro de parcelas no local j .

$\sigma_{adj}^2 = \sigma_{epj}^2 + \sigma_{ej}^2$:= variância ambiental dentro de parcelas no local j .

ρ_a : correlação genética intraclass entre indivíduos da progênie ($\rho_a = 0,25$ para progênies de meios-irmãos e $\rho_a = 0,5$ para progênies de irmãos germanos).

ρ_d : coeficiente de parentesco de dominância entre os indivíduos da progênie ($\rho_d = 0$ para meios-irmãos e $\rho_d = 0,25$ para irmãos germanos).

A análise de variância individual para cada local j , permite a definição dos parâmetros:

$\sigma_{yj}^2 = \sigma_{pj}^2 + \sigma_{ej}^2 + \sigma_{\tilde{g}}^2 + \sigma_{epj}^2$: variância fenotípica do local j .

$h_{aj}^2 = \frac{\sigma_{aj}^2}{\sigma_{yj}^2}$: herdabilidade no sentido restrito no local j .

$h_{gj}^2 = \frac{\sigma_{aj}^2 + \sigma_{dj}^2}{\sigma_{yj}^2}$: herdabilidade no sentido amplo no local j .

$c_j^2 = \frac{\sigma_{ej}^2}{\sigma_{yj}^2}$: correlação intraclass devida ao ambiente comum da parcela no local j .

$$\rho_j = \frac{\sigma_{aj}^2 + \sigma_{d_j}^2 + \sigma_{epj}^2 + \sigma_{ej}^2}{\sigma_{yj}^2} : \text{repetibilidade ao nível de indivíduo no local } j.$$

Análises de variância conjuntas para os locais, tomados dois a dois, permitem a definição dos parâmetros:

σ_a^2 e σ_d^2 : variâncias genética aditiva e de dominância, livres da interação genótipo x ambiente.

$r_a = \frac{\sigma_a^2}{\sigma_{a_1} \sigma_{a_2}}$: correlação genética aditiva entre o desempenho dos indivíduos de um local para outro.

$r_d = \frac{\sigma_d^2}{\sigma_{d_1} \sigma_{d_2}}$: correlação genética de dominância entre o desempenho dos indivíduos de um local para outro.

7.6.1. Seleção por valores genéticos aditivos (propagação sexuada)

No caso do delineamento de blocos casualizados repetidos em vários locais, o método de seleção índice multiefeitos pode ser elaborado da forma apresentada a seguir.

Efeitos aditivos preditos para o local 1, para indivíduos avaliados no local 1

$I_1 = b_1(\bar{Y}_{ilk..} - \bar{Y}_{il..}) + b_2(\bar{Y}_{il..} - \bar{Y}_{.1..}) + b_3(\bar{Y}_{i2..} - \bar{Y}_{.2..}) + b_4(\bar{Y}_{ilk..} - \bar{Y}_{il..} - \bar{Y}_{.lk..} + \bar{Y}_{.1..})$, em que:

$$b_1 = h_{\delta_1}^2 = \frac{(1-\rho_a) \sigma_{a_1}^2}{\sigma_{\delta_1}^2 + \sigma_{e_{11}}^2 / m} = \frac{(1-\rho_a) h_{a1}^2}{[1+(m-1) \rho_1 - mt_{11}] / m}$$

$$b_4 = h_{p_1}^2 = \frac{[(1-\rho_a)/n] \sigma_{a_1}^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2 / n + \sigma_{e_{11}}^2 / nm} = \frac{[1-\rho_a] h_{a1}^2}{[1+(m-1) \rho_1 + (n-1) mt_{11} - nmt_{21}] / m}$$

b_2 e b_3 são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{11} \\ A_{12} \end{bmatrix}, \text{ em que:}$$

$$F_{11} = \sigma_{p_1}^2 + \sigma_{e_1}^2 / b + \sigma_{\delta_1}^2 / (nb) + \sigma_{e_{11}}^2 / (nbt) = \frac{1+(m-1) \rho_1 + (n-1) mt_{11} + (b-1) nmt_{21}}{nbm} \sigma_{F_1}^2$$

$$F_{12} = F_{21} = \rho_a \sigma_a^2 + \rho_d \sigma_d^2 = \rho_a r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2} + \rho_d r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}$$

$$F_{22} = \sigma_{p_2}^2 + \sigma_{e_2}^2 / b + \sigma_{\delta_2}^2 / (nb) + \sigma_{e_{21}}^2 / (nbt) = \frac{1+(m-1) \rho_2 + (n-1) mt_{12} + (b-1) \rho nt_{22}}{nbm} \sigma_{F_2}^2$$

$$A_{11} = \frac{1+(nb-1)}{nb} \rho_a \sigma_{a_1}^2 = \frac{1+(nb-1)}{nb} \rho_a h_{al}^2 \sigma_{y_1}^2$$

$$A_{12} = \rho_a \sigma_a^2 = \rho_a r_a h_{al} h_{a2} \sigma_{y_1} \sigma_{y_2}$$

$$t_{11} = \frac{\rho_a \sigma_{a_1}^2 + \rho_d \sigma_{d_1}^2 + \sigma_{e_1}^2}{\sigma_{y_1}^2} = t_{21} + c_1^2$$

$$t_{21} = \frac{\rho_a \sigma_{a_1}^2 + \rho_d \sigma_{d_1}^2}{\sigma_{y_1}^2} = \rho_a h_{al}^2 + \rho_a (h_{g_1}^2 - h_{al}^2)$$

$$t_{12} = \frac{\rho_a \sigma_{a_2}^2 + \rho_d \sigma_{d_2}^2 + \sigma_{e_2}^2}{\sigma_{y_2}^2} = t_{22} + c_2^2$$

$$t_{22} = \frac{\rho_a \sigma_{a_2}^2 + \rho_d \sigma_{d_2}^2}{\sigma_{y_2}^2} = \rho_a h_{a2}^2 + \rho_d (h_{g_2}^2 - h_{a2}^2)$$

$$D_1 = \left[\frac{\sigma_{d_1}^2}{\sigma_{y_1}^2} \right]^{1/2} = (h_{g_1}^2 - h_{al}^2)^{1/2}$$

$$D_2 = \left[\frac{\sigma_{d_2}^2}{\sigma_{y_2}^2} \right]^{1/2} = (h_{g_2}^2 - h_{a2}^2)^{1/2}$$

$h_{\delta_1}^2$ e $h_{p_1}^2$: herdabilidades dentro de parcela e de parcela, respectivamente, para o local 1.

$\bar{Y}_{i1k..}$, $\bar{Y}_{i1k..}$, $\bar{Y}_{i1...}$, $\bar{Y}_{1k..}$ e $\bar{Y}_{1...}$: valor individual médio no local 1, média da parcela no local 1, média da progénie no local 1, média do bloco no local 1 e média geral no local 1, respectivamente.

$\bar{Y}_{i2...}$ e $\bar{Y}_{2...}$: médias da progénie e geral, respectivamente, no local 2.

Assim, para uma avaliação genética completa com progênies de meios-irmãos ou irmãos germanos são necessários os parâmetros h_a^2 , h_g^2 , c^2 e ρ em cada local e r_a e r_d entre locais. Para progênies de meios-irmãos, quando o objetivo é a predição apenas dos efeitos aditivos, não são necessários os parâmetros h_g^2 e r_d . Para delineamentos com uma planta por parcela tem-se $c^2 = 0$.

Efeitos aditivos preditos para o local 1, para indivíduos avaliados no local 2

$I_2 = b_5 (\bar{Y}_{i2k..} - \bar{Y}_{i2k..}) + b_6 (\bar{Y}_{i1...} - \bar{Y}_{1...}) + b_7 (\bar{Y}_{i2...} - \bar{Y}_{2...}) + b_8 (\bar{Y}_{i2k..} - \bar{Y}_{i2..} - \bar{Y}_{2k..} + \bar{Y}_{2..})$, em que:

$$b_5 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_{\delta_2}^2 + \sigma_{e_2}^2 / m} = \frac{(1-\rho_a) r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_2 - mt_{12}] / m\} \sigma_{y_2}^2}$$

$$b_8 = \frac{[(1-\rho_a)/n] \sigma_a^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2 / n + \sigma_{e_2}^2 / (nm)} = \frac{[(1-\rho_a)/n] r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_2 + (n-1) mt_{12} - nmt_{22}] / m\} \sigma_{y_2}^2}$$

b_6 e b_7 são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_6 \\ b_7 \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{11}^* \\ A_{12}^* \end{bmatrix}, \text{ em que:}$$

$$A_{11}^* = \rho_a \sigma_{a_1}^2 = \rho_a h_{a1}^2 \sigma_{y_1}^2$$

$$A_{12}^* = \frac{1+(nb-1) \rho_a}{nb} \sigma_a^2 = \frac{1+(nb-1) \rho_a}{nb} r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}$$

Efeitos aditivos preditos para o local 2, para indivíduos avaliados no local 1

$I_3 = b_9(\bar{Y}_{ilk\cdot} - \bar{Y}_{ilk..}) + b_{10}(\bar{Y}_{il\cdot\cdot} - \bar{Y}_{l\cdot\cdot\cdot}) + b_{11}(\bar{Y}_{i2\cdot\cdot} - \bar{Y}_{2\cdot\cdot\cdot}) + b_{12}(\bar{Y}_{ilk\cdot} - \bar{Y}_{il\cdot\cdot} - \bar{Y}_{lk\cdot\cdot} + \bar{Y}_{l\cdot\cdot\cdot})$, em que:

$$b_9 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_{\delta_1}^2 + \sigma_{e_1}^2 / m} = \frac{(1-\rho_a) r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_1 - mt_{11}] / m\} \sigma_{y_1}^2}$$

$$b_{12} = \frac{[(1-\rho_a)/n] \sigma_a^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2 / n + \sigma_{e_1}^2 / (nm)} = \frac{[(1-\rho_a)/n] r_a h_{a1} h_{a2} \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_1 + (n-1) mt_{11} - nmt_{21}] / m\} \sigma_{y_2}^2}$$

b_{10} e b_{11} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{10} \\ b_{11} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{21} \\ A_{22} \end{bmatrix}, \text{ em que:}$$

$A_{21} = A_{12}^*$, para o caso de sobrevivências iguais nos dois locais, caso contrário,
 nb refere-se ao número de plantas vivas para a progénie no local 1
para o cômputo de A_{21} , e ao número de plantas vivas para a progê-
nie no local 2 para o cômputo de A_{12}^* .

$$A_{22} = \rho_a \sigma_{a_2}^2 = \rho_a h_{a2}^2 \sigma_{y_2}^2$$

Efeitos aditivos preditos para o local 2, para indivíduos avaliados no local 2

$I_4 = b_{13}(\bar{Y}_{i2kl\cdot} - \bar{Y}_{i2k..}) + b_{14}(\bar{Y}_{il\cdot\cdot} - \bar{Y}_{l\cdot\cdot\cdot}) + b_{15}(\bar{Y}_{i2\cdot\cdot} - \bar{Y}_{2\cdot\cdot\cdot}) + b_{16}(\bar{Y}_{i2k\cdot} - \bar{Y}_{i2\cdot\cdot} - \bar{Y}_{2k\cdot\cdot} + \bar{Y}_{2\cdot\cdot\cdot})$, em que:

$$b_{13} = h_{\delta_2}^2 = \frac{(1-\rho_a) \sigma_{a_2}^2}{\sigma_{\delta_2}^2 + \sigma_{e_2}^2 / m} = \frac{(1-\rho_a) h_{a2}^2}{[1+(m-1) \rho_2 - mt_{12}] / m}$$

$$b_{16} = h_{p_2}^2 = \frac{[(1-\rho_a)/n] \sigma_{a_2}^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2/n + \sigma_{e_{t_2}}^2/(nm)} = \frac{(1-\rho_a) h_{a2}^2}{[1+(m-1)\rho_2 + (n-1)mt_{12} - nmt_{22}]/m}$$

b_{14} e b_{15} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{14} \\ b_{15} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{21}^* \\ A_{22}^* \end{bmatrix}, \text{ em que:}$$

$$A_{21}^* = A_{12}$$

$$A_{22}^* = \frac{1+(nb-1)\rho_a}{nb} \sigma_{a_2}^2 = \frac{1+(nb-1)\rho_a}{nb} h_{a2}^2 \sigma_{y_2}^2$$

No contexto dos modelos lineares mistos, é o seguinte o modelo:

$y = X\beta + Za + Wc + Tp^* + e$, em que:

y, β, a, c, p^* e e : vetores de dados, de efeitos fixos (médias de blocos nos 2 locais), de valores genéticos aditivos (aleatório), de efeitos de parcelas (aleatório), de efeitos permanentes (aleatório) e erros aleatórios (ambiente temporário), respectivamente.

X, Z, W e T : matrizes de incidência para β, a, c e p^* , respectivamente.

Considerando os dois locais, os modelos para cada local são:

$$y_1 = X_1 \beta_1 + Z_1 a_1 + W_1 c_1 + T_1 p_1^* + e_1$$

$$y_2 = X_2 \beta_2 + Z_2 a_2 + W_2 c_2 + T_2 p_2^* + e_2$$

Adotando um modelo multivariado (Henderson & Quaas, 1976; Schaeffer et al., 1978), estes modelos podem ser expressos por:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} p_1^* \\ p_2^* \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Dessa forma, a_1 fornece os valores genéticos preditos de todos os indivíduos para o local 1 e a_2 , os valores genéticos preditos de todos os indivíduos para o local 2.

Para o modelo bivariado, é a seguinte a estrutura de variâncias:

$$Var = \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1^* \\ p_2^* \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \sigma_{a_1}^2 A & \sigma_a^2 A & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_a^2 A & \sigma_{a_2}^2 A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{e_1}^2 I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{e_2}^2 I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{ep_1}^2 I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{ep_2}^2 I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{et_1}^2 I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{et_2}^2 I \end{bmatrix}$$

O modelo bivariado considera, então, o mesmo caráter avaliado em diferentes indivíduos, nos diferentes locais, como sendo caracteres diferentes. Entretanto, em função da avaliação a ser realizada em indivíduos diferentes, não existem covariâncias entre os efeitos de parcela nos dois locais, entre os efeitos permanentes nos dois locais e entre os efeitos residuais dentro de parcela nos dois locais.

As equações de modelo misto para predição dos valores genéticos pelo procedimento BLUP sob modelo individual (que devem fornecer, de modo simultâneo, os mesmos resultados dados pelo vetor formado pelos índices I_1 , I_2 , I_3 e I_4 , em situações de dados balanceados) equivalem a:

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

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

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

$$R_O = \begin{bmatrix} \sigma_{et_1}^2 & 0 \\ 0 & \sigma_{et_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad P_O^* = \begin{bmatrix} \sigma_{ep_1}^2 & 0 \\ 0 & \sigma_{ep_2}^2 \end{bmatrix}$$

\otimes : denota a operação produto de Kronecker.

Os componentes de variância podem ser obtidos a partir dos parâmetros h_a^2 , ρ , c^2 , σ_y^2 e r_a , da seguinte maneira:

$$\begin{aligned}\sigma_{aj}^2 &= h_{aj}^2 \sigma_{yj}^2; & \sigma_{ej}^2 &= c_j^2 \sigma_{yj}^2; & \sigma_{ep^*j}^2 &= (\sigma_{epj}^2 + \sigma_{dj}^2) = (\rho_j - h_{aj}^2 - c_j^2) \sigma_{yj}^2; \\ \sigma_{etj}^2 &= (1 - \rho_j - c_j^2) \sigma_{yj}^2; & \sigma_a^2 &= r_a \sigma_{a_1} \sigma_{a_2}.\end{aligned}$$

Pelo modelo e equações apresentados, o vetor p^* de efeitos permanentes contempla os efeitos de ambiente permanente somados aos efeitos de dominância. Assim, este modelo considera que toda a variação de dominância encontra-se dentro de progênies, o que é incorreto para progênies de irmãos gêmeos.

7.6.2. Seleção por valores genotípicos (propagação assexuada)

Para a predição dos efeitos de dominância, o método índice multiefeitos pode ser estruturado conforme relatado a seguir.

Efeitos de dominância preditos para o local 1, para indivíduos avaliados no local 1

$I_5 = b_{17}(\bar{Y}_{ilk..} - \bar{Y}_{ilk..}) + b_{18}(\bar{Y}_{il...} - \bar{Y}_{1...}) + b_{19}(\bar{Y}_{i2...} - \bar{Y}_{2...}) + b_{20}(\bar{Y}_{ilk..} - \bar{Y}_{il...} - \bar{Y}_{1k..} + \bar{Y}_{1...})$, em que:

$$b_{17} = \frac{(1 - \rho_d) \sigma_{d_1}^2}{\sigma_{\delta_1}^2 + \sigma_{e_1}^2 / m} = \frac{(1 - \rho_d) (h_{g_1}^2 - h_{a_1}^2)}{[1 + (m-1) \rho_1 - mt_{11}] / m};$$

$$b_{20} = \frac{[(1 - \rho_d) / n] \sigma_{d_1}^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2 / n + \sigma_{e_1}^2 / (nm)} = \frac{(1 - \rho_d) (h_{g_1}^2 - h_{a_1}^2)}{[1 + (m-1) \rho_1 + (n-1) mt_{11} - nm t_{21}] / m}$$

$$\begin{bmatrix} b_{18} \\ b_{19} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{11} \\ D_{12} \end{bmatrix}, \text{ em que:}$$

$$D_{11} = \frac{1 + (nb-1) \rho_d}{nb} \sigma_{d_1}^2 = \frac{1 + (nb-1) \rho_d}{nb} (h_{g_1}^2 - h_{a_1}^2) \sigma_{y_1}^2$$

$$D_{12} = \rho_d \sigma_d^2 = \rho_d r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}$$

Efeitos de dominância preditos para o local 1, para indivíduos avaliados no local 2

$I_6 = b_{21}(\bar{Y}_{i2k\ell} - \bar{Y}_{i2k..}) + b_{22}(\bar{Y}_{il..} - \bar{Y}_{l...}) + b_{23}(\bar{Y}_{i2..} - \bar{Y}_{2...}) + b_{24}(\bar{Y}_{i2k..} - \bar{Y}_{i2...} - \bar{Y}_{2k..} + \bar{Y}_{2...})$, em que:

$$b_{21} = \frac{(1-\rho_d) \sigma_d^2}{\sigma_{\delta_2}^2 + \sigma_{e_{t2}}^2 / m} = \frac{(1-\rho_d) r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_2 - mt_{12}] / m\} \sigma_{y_2}^2}$$

$$b_{24} = \frac{[(1-\rho_d)/n] \sigma_d^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2 / n + \sigma_{e_{t2}}^2 / (nm)} = \frac{[(1-\rho_d)/n] r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_2 + (n-1) mt_{12} - nmt_{22}] / m\} \sigma_{y_2}^2}$$

b_{22} e b_{23} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{22} \\ b_{23} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{11}^* \\ D_{12}^* \end{bmatrix}, \text{ em que:}$$

$$D_{11}^* = \rho_d \sigma_{d_1}^2 = \rho_d (h_{g_1}^2 - h_{a1}^2) \sigma_{y_1}^2$$

$$D_{12}^* = \frac{1+(nb-1)}{nb} \rho_d \sigma_d^2 = \frac{1+(nb-1)}{nb} \rho_d r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}$$

Efeitos de dominância preditos para o local 2, para indivíduos avaliados no local 1

$I_7 = b_{25}(\bar{Y}_{ilk\ell} - \bar{Y}_{ilk..}) + b_{26}(\bar{Y}_{il..} - \bar{Y}_{l...}) + b_{27}(\bar{Y}_{i2..} - \bar{Y}_{2...}) + b_{28}(\bar{Y}_{ilk..} - \bar{Y}_{il..} - \bar{Y}_{lk..} + \bar{Y}_{l...})$, em que:

$$b_{25} = \frac{(1-\rho_d) \sigma_d^2}{\sigma_{\delta_1}^2 + \sigma_{e_{t1}}^2 / m} = \frac{(1-\rho_d) r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_1 - mt_{11}] / m\} \sigma_{y_1}^2}$$

$$b_{28} = \frac{[(1-\rho_d)/n] \sigma_d^2}{\sigma_{e_1}^2 + \sigma_{\delta_1}^2 / n + \sigma_{e_{t1}}^2 / nm} = \frac{[(1-\rho_d)/n] r_d D_1 D_2 \sigma_{y_1} \sigma_{y_2}}{\{[1+(m-1) \rho_1 + (n-1) mt_{11} - nmt_{21}] / m\} \sigma_{y_2}^2}$$

b_{26} e b_{27} são obtidos por meio do sistema:

$$\begin{bmatrix} b_{26} \\ b_{27} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{21} \\ D_{22} \end{bmatrix}, \text{ em que:}$$

$D_{21} = D_{12}^*$, para o caso de sobrevivências iguais nos dois locais.

$$D_{22} = \rho_d \sigma_{d_2}^2 = \rho_d (h_{g_2}^2 - h_{a2}^2) \sigma_{y_2}^2$$

Efeitos de dominância preditos para o local 2, para indivíduos avaliados no local 2

$I_8 = b_{29}(\bar{Y}_{i2k\ell} - \bar{Y}_{i2k..}) + b_{30}(\bar{Y}_{il..} - \bar{Y}_{l...}) + b_{31}(\bar{Y}_{i2..} - \bar{Y}_{2...}) + b_{32}(\bar{Y}_{i2k..} - \bar{Y}_{i2...} - \bar{Y}_{2k..} + \bar{Y}_{2...})$, em que:

$$b_{29} = \frac{(1-\rho_d) \sigma_{d_2}^2}{\sigma_{\delta_2}^2 + \sigma_{e_{t_2}}^2 / m} = \frac{(1-\rho_d) (h_{g_2}^2 - h_{a2}^2)}{[1+(m-1) \rho_2 - mt_{12}] / m}$$

$$b_{32} = \frac{[(1-\rho_d) / n] \sigma_{d_2}^2}{\sigma_{e_2}^2 + \sigma_{\delta_2}^2 / n + \sigma_{e_{t_2}}^2 / (nm)} = \frac{(1-\rho_d) (h_{g_2}^2 - h_{a2}^2)}{[1+(m-1) \rho_2 + (n-1) mt_{12} - nmt_{22}] / m}$$

b_{30} e b_{31} são obtidos resolvendo-se o sistema:

$$\begin{bmatrix} b_{30} \\ b_{31} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{21}^* \\ D_{22}^* \end{bmatrix}, \text{ em que:}$$

$$D_{21}^* = D_{12}$$

$$D_{22}^* = \frac{1+(nb-1) \rho_d}{nb} \sigma_{d_2}^2 = \frac{1+(nb-1) \rho_d}{nb} (h_{g_2}^2 - h_{a2}^2) \sigma_{y_2}^2$$

No contexto dos modelos lineares mistos, adotando-se o modelo multivariado, tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

em que d_1 e d_2 referem-se aos vetores dos efeitos de dominância no locais 1 e 2, respectivamente, e p_1 e p_2 , aos efeitos de ambiente permanente nos locais 1 e 2, respectivamente.

As equações de modelo misto para a predição dos efeitos aditivos e de dominância pelo procedimento BLUP sob modelo individual (que devem fornecer os mesmos resultados dados pelo vetor formado pelos índices $I_1, I_2, I_3, I_4, I_5, I_6, I_7$ e I_8 , em situações de dados balanceados) equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{d} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}T \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} & Z'R^{-1}Z & Z'R^{-1}W & Z'R^{-1}T \\ Z'R^{-1}X & Z'R^{-1}Z & Z'R^{-1}Z+G_d^{-1} & Z'R^{-1}W & Z'R^{-1}T \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}Z & W'R^{-1}W+C^{-1} & W'R^{-1}T \\ T'R^{-1}X & T'R^{-1}Z & T'R^{-1}Z & T'R^{-1}W & T'R^{-1}T+P^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ T'R^{-1}y \end{bmatrix} \quad (2), \text{ em que:}$$

$$\hat{d} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \end{bmatrix}; \quad \hat{p} = \begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \end{bmatrix}$$

$$G_d^{-1} = G_{do}^{-1} \otimes D^{-1}; \quad P^{-1} = P_o^{-1} \otimes I$$

D : matriz de parentesco de dominância entre os indivíduos em avaliação.

$$G_{do} = \begin{bmatrix} \sigma_{d_1}^2 & \sigma_d^2 \\ \sigma_d^2 & \sigma_{d_2}^2 \end{bmatrix}; \quad P_o = \begin{bmatrix} \sigma_{ep_1}^2 & 0 \\ 0 & \sigma_{ep_2}^2 \end{bmatrix}$$

Os componentes de variância podem ser obtidos a partir dos parâmetros h_a^2 , h_g^2 , ρ , e r_d por meio de:

$$\sigma_{d_j}^2 = (h_{gj}^2 - h_{aj}^2) \sigma_{yj}^2; \quad \sigma_d^2 = r_d \sigma_{d_1} \sigma_{d_2}; \quad \sigma_{epj}^2 = (\rho_j - h_{gj}^2 - c_j^2) \sigma_{yj}^2.$$

Os efeitos genotípicos preditos dos indivíduos são dados por: $\hat{g} = \hat{a} + \hat{d}$ e, sendo $Var(g) = G_o \otimes A + G_{do} \otimes D = G_{ad}$, os valores genotípicos podem ser preditos diretamente por:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}T \\ Z'R^{-1}X & Z'R^{-1}Z + G_{ad}^{-1} & Z'R^{-1}W & Z'R^{-1}T \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}T \\ T'R^{-1}X & T'R^{-1}Z & T'R^{-1}W & T'R^{-1}T + P^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ T'R^{-1}y \end{bmatrix} \quad (3)$$

Os efeitos genotípicos preditos diretamente pelo índice multiefeitos são obtidos conforme relatado a seguir:

- (a) Efeitos genotípicos preditos para o local 1, para indivíduos avaliados no local 1.
 $I_9 = I_1 + I_5$
- (b) Efeitos genotípicos preditos para o local 1, para indivíduos avaliados no local 2.
 $I_{10} = I_2 + I_6$
- (c) Efeitos genotípicos preditos para o local 2, para indivíduos avaliados no local 1.
 $I_{11} = I_3 + I_7$
- (d) Efeitos genotípicos preditos para o local 2, para indivíduos avaliados no local 2.
 $I_{12} = I_4 + I_8$

Os índices I_9 , I_{10} , I_{11} e I_{12} podem ser aplicados à seleção de clones em testes clonais, bastando considerar $\rho_a = \rho_d = 1$.

7.6.3. Exemplo Prático

Consideraram-se, para exercício das metodologias, os seguintes componentes de variância:

Local 1:

$$h_{a1}^2 = 0,1457; \quad h_{g1}^2 = 0,2185; \quad c_1^2 = 0,1166; \quad \rho_1 = 0,4166; \quad \sigma_{y1}^2 = 7,4747.$$

$$\sigma_{p1}^2 = 0,2722; \quad \sigma_{e1}^2 = 0,8716; \quad \sigma_{\delta_1}^2 = 1,97; \quad \sigma_{ep1}^2 = 0,6092; \quad \sigma_{ep1}^{2*} = 1,1536; \quad \sigma_{et1}^2 = 4,3610;$$

$$\sigma_{a1}^2 = 1,0888; \quad \sigma_{d1}^2 = 0,5444$$

Local 2:

$$h_{a2}^2 = 0,1875; \quad h_{g2}^2 = 0,28125; \quad c_2^2 = 0,13; \quad \rho_2 = 0,48; \quad \sigma_{y2}^2 = 8,0.$$

$$\sigma_{p2}^2 = 0,375; \quad \sigma_{e2}^2 = 1,04; \quad \sigma_{\delta_2}^2 = 2,425; \quad \sigma_{ep2}^2 = 0,55; \quad \sigma_{ep2}^{2*} = 1,3; \quad \sigma_{et2}^2 = 4,16; \quad \sigma_{a2}^2 = 1,5; \quad \sigma_{d2}^2 = 0,75$$

Análise conjunta dos locais 1 e 2:

$$r_a = 0,7; \quad \sigma_a^2 = 0,8946$$

$$r_d = 0,8; \quad \sigma_d^2 = 0,5112$$

Para aplicação das metodologias, considerou-se o seguinte subconjunto de dados associados a um teste de progêneres de meios-irmãos ($\rho_a = 0,25$ e $\rho_d = 0$) de *Eucalyptus grandis*.

Obs.	Local	Indi- víduo	Bloco	Pro- gênie	Indivíduos/ Progênie/ Local	Medição	Parcela	Diâmetro (cm)
1	1	1	1	1	1	1	1	9,87
2	1	1	1	1	1	2	1	10,00
3	1	2	1	1	2	1	1	14,48
4	1	2	1	1	2	2	1	15,00
5	1	3	1	2	1	1	2	8,91
6	1	3	1	2	1	2	2	9,00
7	1	4	1	2	2	1	2	15,00
8	1	4	1	2	2	2	2	14,64
9	1	5	2	1	1	1	3	9,55
10	1	5	2	1	1	2	3	10,00
11	1	6	2	1	2	1	3	8,00
12	1	6	2	1	2	2	3	7,96
13	1	7	2	2	1	1	4	15,00
14	1	7	2	2	1	2	4	16,07
15	1	8	2	2	2	1	4	14,01
16	1	8	2	2	2	2	4	14,50
17	2	9	3	1	1	1	5	10,82
18	2	9	3	1	1	2	5	12,10
19	2	10	3	1	2	1	5	12,10
20	2	10	3	1	2	2	5	11,78
21	2	11	3	2	1	1	6	11,46
22	2	11	3	2	1	2	6	11,78
23	2	12	3	2	2	1	6	13,69
24	2	12	3	2	2	2	6	7,32
25	2	13	4	1	1	1	7	9,87
26	2	13	4	1	1	2	7	9,87
27	2	14	4	1	2	1	7	9,39
28	2	14	4	1	2	2	7	13,85
29	2	15	4	2	1	1	8	9,71
30	2	15	4	2	1	2	8	15,12
31	2	16	4	2'	2	1	8	14,80
32	2	16	4	2	2	2	8	7,32

Os vetores e matrizes apresentam as seguintes dimensões neste exemplo:

$$y = (64 \times 1); \beta = (4 \times 1); a = d = (36 \times 1) \text{ incluindo as duas matrizes (parentais)}; \\ c = (8 \times 1); p^* = p = (16 \times 1); X = (64 \times 4); Z = (64 \times 36); W = (64 \times 8); T = (64 \times 16); \\ A = (18 \times 18); D = I_{(18)}; R^{-1} = R_o^{-1} \otimes I_{(32)}; C^{-1} = C_o^{-1} \otimes I_{(4)}; P^{-1} = P_o^{-1} \otimes I_{(8)}.$$

É importante ressaltar que o vetor y apresenta seqüencialmente: 16 valores de diâmetro dos indivíduos do local 1, 32 valores zero e 16 valores de diâmetro dos indivíduos do local 2.

Seleção por valores genéticos aditivos

Os resultados referentes ao método BLUP (1), os quais são idênticos aos obtidos pelos índices I_1 a I_4 , são apresentados na Tabela 7.13.

Tabela 7.13. Coeficientes de ponderação (b_i) dos diferentes efeitos do modelo estatístico pelos métodos I_1 a I_4 e efeitos genéticos aditivos preditos por estes índices pelo método BLUP (1).

b_1	Indivíduos	Efeitos genéticos preditos para o local 1	Efeitos genéticos preditos para o local 2
$b_1 = 0,19674$	1	-0,6211	-0,5018
$b_2 = 0,2624$	2	0,3242	0,2750
$b_3 = 0,0816$	3	-0,4285	-0,3606
$b_4 = 0,13855$	4	0,7254	0,5874
$b_5 = 0,14893$	5	-0,4199	-0,3364
$b_6 = 0,1330$	6	-0,7731	-0,6266
$b_7 = 0,1789$	7	0,7224	0,5850
$b_8 = 0,10189$	8	0,4706	0,3781
$b_9 = 0,16165$	9	-0,1952	-0,1404
$b_{10} = 0,2033$	10	-0,1237	-0,0206
$b_{11} = 0,1630$	11	0,2425	0,2197
$b_{12} = 0,11384$	12	0,0765	-0,0587
$b_{13} = 0,244972$	13	-0,3728	-0,4382
$b_{14} = 0,0878$	14	-0,1122	-0,0012
$b_{15} = 0,3150$	15	0,3434	0,3889
$b_{16} = 0,1708$	16	0,1416	0,0505
	17*	-0,4171	-0,3255
	18**	0,4171	0,3255

* e ** - correspondem às matrizes 1 e 2, respectivamente.

Verifica-se que um mesmo indivíduo apresenta dois valores genéticos preditos, um para cada local em que for utilizado (Tabela 7.13). Isto permite maior eficiência dos programas de melhoramento em presença de interação genótipo x ambiente. Para o caso de dados balanceados, verifica-se a equivalência entre os métodos índice multiefeitos e BLUP.

Seleção por valores genotípicos

Os resultados referentes aos efeitos de dominância preditos pelo método BLUP (2), e índices I_5 a I_8 , são apresentados na Tabela 7.14.

Tabela 7.14. Coeficientes de ponderação (b_i) dos diferentes efeitos do modelo estatístico pelos métodos I_5 a I_8 e efeitos de dominância preditos por estes índices e pelo método BLUP(2).

b_1	Indivíduos	Efeitos de dominância preditos para o local 1	Efeitos de dominância preditos para o local 2
$b_{17} = 0,1312$	1	-0,2751	-0,2583
$b_{18} = 0,0791$	2	0,3552	0,3335
$b_{19} = -0,0088$	3	-0,4247	-0,3988
$b_{20} = 0,0924$	4	0,3446	0,3236
$b_{21} = 0,11350$	5	-0,1409	-0,1323
$b_{22} = -0,0082$	6	-0,3764	-0,3534
$b_{23} = 0,0641$	7	0,3426	0,3217
$b_{24} = 0,0776$	8	0,1747	0,1640
$b_{25} = 0,1232$	9	0,0101	0,0149
$b_{26} = 0,0743$	10	0,0646	0,0948
$b_{27} = -0,0082$	11	0,0259	0,0380
$b_{28} = 0,0867$	12	-0,1006	-0,1477
$b_{29} = 0,1665$	13	-0,1252	-0,1837
$b_{30} = -0,0121$	14	0,0734	0,1077
$b_{31} = 0,0941$	15	0,1028	0,1508
$b_{32} = 0,1139$	16	-0,0510	-0,0748
	17*	0,0000	0,0000
	18**	0,0000	0,0000

* e ** - correspondem às matrizes 1 e 2, respectivamente.

Os demais efeitos estimados e preditos pelo BLUP (2), exceto os efeitos aditivos (que foram apresentados na Tabela 7.14), são apresentados na Tabela 7.15.

Tabela 7.15. Efeitos fixos de blocos ($\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$) e efeitos de parcela (\hat{c}) e permanentes (\hat{p}) estimados e preditos pelo método BLUP (1).

$\hat{\beta}$	\hat{c}	\hat{p}
$\hat{\beta}_1 = 12,1125$	$\hat{c}_1 = -0,1282$	$\hat{p}_1 = -0,3078$
$\hat{\beta}_2 = 11,8863$	$\hat{c}_2 = -0,1282$	$\hat{p}_2 = 0,3974$
$\hat{\beta}_3 = 11,3813$	$\hat{c}_3 = -0,8282$	$\hat{p}_3 = -0,4752$
$\hat{\beta}_4 = 11,2413$	$\hat{c}_4 = 0,8282$	$\hat{p}_4 = 0,3856$
	$\hat{c}_5 = 0,1521$	$\hat{p}_5 = -0,1577$
	$\hat{c}_6 = -0,1521$	$\hat{p}_6 = -0,4212$
	$\hat{c}_7 = -0,1053$	$\hat{p}_7 = 0,3834$
	$\hat{c}_8 = 0,1053$	$\hat{p}_8 = 0,1955$
		$\hat{p}_9 = 0,0109$
		$\hat{p}_{10} = 0,0695$
		$\hat{p}_{11} = 0,0278$
		$\hat{p}_{12} = -0,1083$
		$\hat{p}_{13} = -0,1347$
		$\hat{p}_{14} = 0,0790$
		$\hat{p}_{15} = 0,1106$
		$\hat{p}_{16} = -0,0549$

Sendo os efeitos de blocos hierárquicos em relação ao efeito de locais, não há necessidade de se incluir os efeitos de locais em β , já que β , neste caso, fornece a estimação para todos os blocos ao longo dos locais.

Na Tabela 7.16. são apresentados os efeitos genotípicos preditos pelos métodos I_9 a I_{12} e BLUP (3).

Tabela 7.16. Efeitos genotípicos ($\hat{g} = \hat{a} + \hat{d}$) preditos pelos procedimentos I_9 a I_{12} e BLUP(3).

Indivíduos	\hat{g} no local 1	\hat{g} no local 2
1	-0,8962	-0,7600
2	0,6793	0,6085
3	-0,8531	-0,7594
4	1,0700	0,9110
5	-0,5609	-0,4688
6	-1,1494	-0,9800
7	1,0650	0,9067
8	0,6453	0,5421
9	-0,1851	-0,1255
10	-0,0591	0,0742
11	0,2684	0,2577
12	-0,0242	-0,2064
13	-0,4980	-0,6219
14	-0,0388	0,1065
15	0,4462	0,5397
16	0,0906	-0,0243
17	-0,4171	-0,3255
18	0,4171	0,3255

Os resultados obtidos revelaram uma grande eficiência e flexibilidade do procedimento BLUP individual sob modelo multivariado para a predição de valores genéticos e genotípicos em presença de interação genótipo x ambiente e medidas repetidas.

Na seleção que envolver indivíduos de diferentes gerações, o modelo multivariado também é o mais adequado, sendo que os experimentos de diferentes gerações (gerações subsequentes, por exemplo) podem ser analisados simultaneamente considerando diferentes gerações como se fossem diferentes locais.

Considerações Finais

- ii) As equações de modelo misto apresentadas, associadas ao BLUP individual multivariado, permitem a predição de valores genéticos e genotípicos para cada indivíduo em cada local em que for utilizado, considerando de maneira eficiente a ocorrência de interação genótipo x ambiente e a presença de medidas repetidas; (ii) para pleno uso destas equações de modelo misto, são necessárias estimativas dos parâmetros herdabilidade no sentido restrito, herdabilidade no sentido amplo, repetibilidade, correlação devida ao ambiente comum em cada local e correlação genética aditiva e de dominância através dos locais.

7.7. BLUP Individual e Índice Multiefeitos Associados a Cru-

zamentos Fatoriais e Dialélicos

O uso de elaborados delineamentos de cruzamento tem sido prática enfatizada em gerações avançadas de programas de melhoramento florestal (Burdon & Shelbourne, 1971; Van Buijtenen & Burdon, 1990). Dentre estes delineamentos, destacam-se os do tipo dialélico parcial, fatorial desconexo e meio dialélico desconexo. Variações destes delineamentos têm sido muito utilizadas no melhoramento de *Pinus* (Cotterill, 1986; Cotterill et al., 1987) e *Eucalyptus* (Bouvet & Vigneron, 1995). A seguir são apresentados os desenvolvimentos realizados por Resende (1999b).

7.7.1. Modelo Estatístico

O modelo estatístico associado à avaliação (no delineamento de blocos casualizados) de progêniés de irmãos germanos obtidas sob cruzamentos fatoriais equivale a:

$$Y_{ijkl} = \mu + f_i + m_j + b_k + fm_{ij} + e_{ijk} + \delta_{ijkl}, \text{ em que:}$$

μ : efeito da média geral, fixo, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$;

f_i : efeito do genitor feminino i , aleatório, $E(f_i) = 0$ e $E(f_i^2) = \sigma_f^2$;

m_j : efeito do genitor masculino j , aleatório, $E(m_j) = 0$ e $E(m_j^2) = \sigma_m^2$;

b_k : efeito do bloco k , fixo, $E(b_k) = b_k$ e $E(b_k^2) = V_b$;

fm_{ij} : efeito da progénie oriunda do cruzamento entre os genitores i e j , aleatório, $E(fm_{ij}) = 0$ e $E(fm_{ij}^2) = \sigma_{fm}^2$;

e_{ijk} : efeito da parcela ijk , aleatório, $E(e_{ijk}) = 0$ e $E(e_{ijk}^2) = \sigma_e^2$;

δ_{ijkl} : efeito do indivíduo ℓ dentro da parcela ijk , aleatório, $E(\delta_{ijkl}) = 0$ e $E(\delta_{ijkl}^2) = \sigma_\delta^2$.

O modelo pode ser reescrito da seguinte forma:

$$Y_{ijkl} = \bar{Y}_{...} + (\bar{Y}_{i...} - \bar{Y}_{...}) + (\bar{Y}_{j..} - \bar{Y}_{...}) + (\bar{Y}_{..k} - \bar{Y}_{...}) + (\bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{j..} + \bar{Y}_{...}) + \\ + (\bar{Y}_{jk..} - \bar{Y}_{ij..} - \bar{Y}_{..k} + \bar{Y}_{...}) + (Y_{ijkl} - \bar{Y}_{ijk.}),$$

em que:

$\bar{Y}_{...}$: média geral.

$(\bar{Y}_{i...} - \bar{Y}_{...})$: efeito da capacidade geral de combinação do genitor i .

$(\bar{Y}_{j..} - \bar{Y}_{...})$: efeito da capacidade geral de combinação do genitor j .

$(\bar{Y}_{..k} - \bar{Y}_{...})$: efeito de bloco.

$(\bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{j..} + \bar{Y}_{....})$: efeito da capacidade específica de combinação entre os genitores i e j .

$(\bar{Y}_{ijk.} - \bar{Y}_{ij..} - \bar{Y}_{..k.} + \bar{Y}_{....})$: efeito de parcela.

$(Y_{ijkl} - \bar{Y}_{ijk.})$: efeito de indivíduo dentro de parcela.

7.7.2. Índice multiefeitos para os efeitos aditivos

Nesta situação, o índice multiefeitos para predição de efeitos aditivos (a) é dado por:

$$I_1 = b_1 f_i + b_2 m_j + b_3 fm_j + b_4 e_{ijk} + b_5 \delta_{ijkl}$$

Os pesos b_i devem ser determinados de forma que se maximize a correlação entre o índice e os efeitos genéticos a serem preditos, empregando-se a técnica da covariância entre parentes, conforme Resende (1991) e Resende & Higa (1994). Assim, têm-se as seguintes covariâncias e variâncias:

$$V_1 = Var(\bar{Y}_{i...} - \bar{Y}_{....}) = \frac{f-1}{f} [\sigma_f^2 + \sigma_{fm}^2 / m + \sigma_e^2 / (mb) + \sigma_\delta^2 / (mnb)]$$

$$V_2 = Var(\bar{Y}_{j..} - \bar{Y}_{....}) = \frac{m-1}{m} [\sigma_m^2 + \sigma_{fm}^2 / f + \sigma_e^2 / (fb) + \sigma_\delta^2 / (fnb)]$$

$$V_3 = Var(\bar{Y}_{ij..} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{....}) = \frac{f-1}{f} \frac{m-1}{m} [\sigma_{fm}^2 + \sigma_e^2 / b + \sigma_\delta^2 / (nb)]$$

$$V_4 = Var(\bar{Y}_{ijk.} - \bar{Y}_{ij..} - \bar{Y}_{..k.} + \bar{Y}_{....}) = \left(\frac{fm-1}{fm} \right) \left(\frac{b-1}{b} \right) (\sigma_e^2 + \sigma_\delta^2 / n)$$

$$V_5 = Var(Y_{ijkl} - \bar{Y}_{ijk.}) = \frac{n-1}{n} \sigma_\delta^2$$

As equivalências em termos de componentes de variância são:

$$\sigma_f^2 = \sigma_m^2 = (1/4) \sigma_a^2$$

$$\sigma_{fm}^2 = (1/4) \sigma_d^2$$

$$\sigma_y^2 = \sigma_f^2 + \sigma_m^2 + \sigma_{fm}^2 + \sigma_e^2 + \sigma_\delta^2$$

Assim, as variâncias V_1 a V_5 podem ser expressas alternativamente por:

$$V_1 = \frac{f-1}{f} \left[\frac{1 + (n-1)t_1 + (b-1)nt_2 + (m-2)nbt_3}{nbtm} \right] \sigma_y^2$$

$$V_2 = \frac{m-1}{m} \left[\frac{1+(n-1) t_1 + (b-1) n t_2 + (f-2) n b t_3}{n b f} \right] \sigma_y^2$$

$$V_3 = \frac{f-1}{f} \frac{m-1}{m} \left[\frac{1+(n-1) t_1 + (b-1) n t_2 - 2 n b t_3}{n b} \right] \sigma_y^2$$

$$V_4 = \frac{f m - 1}{f m} \left[\frac{1+(n-1) t_1 - n t_2}{n} \right] \sigma_y^2$$

$$V_5 = \frac{n-1}{n} (1-t_1) \sigma_y^2, \text{ em que:}$$

$t_1 = \frac{(1/2) \sigma_a^2 + (1/4) \sigma_d^2 + \sigma_e^2}{\sigma_y^2} = (1/2) h_a^2 + (1/4) (h_g^2 - h_a^2) + c^2 = t_2 + c^2$: correlação intraclasse entre os indivíduos dentro da parcela.

$t_2 = \frac{(1/2) \sigma_a^2 + (1/4) \sigma_d^2}{\sigma_y^2} = (1/2) h_a^2 + (1/4) (h_g^2 - h_a^2)$: correlação intraclasse entre os indivíduos da progénie de irmãos germanos em diferentes parcelas.

$t_3 = \frac{(1/2) \sigma_a^2}{\sigma_y^2} = (1/4) h_a^2$: correlação intraclasse entre os indivíduos meios-irmãos em diferentes parcelas.

$c^2 = \frac{\sigma_e^2}{\sigma_y^2}$: correlação entre os indivíduos na parcela, devida ao ambiente comum da parcela.

$h_a^2 = \frac{\sigma_a^2}{\sigma_y^2}$: herdabilidade no sentido restrito.

$h_g^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_y^2}$: herdabilidade no sentido amplo.

σ_a^2 : variância genética aditiva.

σ_d^2 : variância genética de dominância.

σ_y^2 : variância fenotípica.

Têm-se, também, as seguintes covariâncias entre efeitos aditivos (a) e valores fenotípicos:

$$COV(a, \bar{Y}_{i...}) = 2 \left[\frac{1}{nbm} COV(M, F) + \frac{nb(m-1)}{nbm} COV(MTS) + \frac{nb-1}{nbm} COV(TS) \right]$$

$$= \frac{\sigma_a^2}{nbm} + \frac{nb(m-1)}{nbm} (1/4) \sigma_a^2 + \frac{nb-1}{nbm} (1/2) \sigma_a^2$$

$$COV(a, \bar{Y}_{...}) = 2 \left[\frac{1}{nbfm} COV(M, F) + \frac{nb(m-1)}{nbfm} COV(MTS) + \frac{nb(f-1)}{nbfm} COV(MTS) + \frac{nb-1}{nbfm} COV(TS) \right]$$

$$= \frac{\sigma_a^2}{nbfm} + \frac{nb(m-1)}{nbfm} (1/4) \sigma_a^2 + \frac{nb(f-1)}{nbfm} (1/4) \sigma_a^2 + \frac{nb-1}{nbfm} (1/2) \sigma_a^2$$

$$COV(a, \bar{Y}_{j...}) = 2 \left[\frac{1}{nbf} COV(M, F) + \frac{nb(f-1)}{nbf} COV(MTS) + \frac{nb-1}{nbf} COV(TS) \right]$$

$$= \frac{\sigma_a^2}{nbf} + \frac{nb(f-1)}{nbf} (1/4) \sigma_a^2 + \frac{nb-1}{nbf} (1/2) \sigma_a^2$$

$$COV(a, \bar{Y}_{...k}) = 2 \left[\frac{1}{nfm} COV(M, F) + \frac{n(m-1)}{nfm} COV(MTS) + \frac{n(f-1)}{nfm} COV(MTS) + \frac{n-1}{nfm} COV(TS) \right]$$

$$= \frac{\sigma_a^2}{nfm} + \frac{n(m-1)}{nfm} (1/4) \sigma_a^2 + \frac{n(f-1)}{nfm} (1/4) \sigma_a^2 + \frac{n-1}{nfm} (1/2) \sigma_a^2$$

$$COV(a, \bar{Y}_{ij...}) = 2 \left[\frac{1}{nb} COV(M, F) + \frac{nb-1}{nb} COV(TS) \right]$$

$$= \frac{\sigma_a^2}{nb} + \frac{nb-1}{nb} (1/2) \sigma_a^2$$

$$COV(a, \bar{Y}_{ijk...}) = 2 \left[\frac{1}{n} COV(M, F) + \frac{n-1}{n} COV(TS) \right]$$

$$= \frac{\sigma_a^2}{n} + \frac{n-1}{n} (1/2) \sigma_a^2$$

$$COV(a, \bar{Y}_{ijkl}) = 2 COV(M, F), \text{ em que:}$$

$$= \sigma_a^2$$

$COV(M, F)$, $COV(MTS)$ e $COV(TS)$: covariâncias mãe-filho, meio-tio-sobrinho e tio-sobrinho, respectivamente.

Com base nestas covariâncias, obtém-se as seguintes covariâncias para os efeitos:

a) Efeito da capacidade geral de combinação do genitor i (feminino)

$$COV[(a, (\bar{Y}_{i...} - \bar{Y}_{...}))] = \frac{f-1}{f} \left[(1/4) \sigma_a^2 + \frac{(1/2) \sigma_a^2}{nbm} \right]$$

b) Efeito da capacidade geral de combinação do genitor j (masculino)

$$COV[(a, (\bar{Y}_{j..} - \bar{Y}_{...}))] = \frac{m-1}{m} \left[(1/4) \sigma_a^2 + \frac{(1/2) \sigma_a^2}{nbf} \right]$$

c) Efeito da capacidade específica de combinação do genitor i com o j

$$COV[(a, (\bar{Y}_{ij..} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}))] = \frac{f-1}{f} \frac{m-1}{m} \frac{(1/2) \sigma_a^2}{nb}$$

d) Efeito da parcela ijk

$$COV[(a, (\bar{Y}_{ijk..} - \bar{Y}_{ij..} - \bar{Y}_{ik..} + \bar{Y}_{...}))] = \frac{fm-1}{fm} \frac{b-1}{b} \frac{(1/2) \sigma_a^2}{n}$$

e) Efeito do indivíduo na parcela.

$$COV[(a, (\bar{Y}_{ijkl..} - \bar{Y}_{ijk..}))] = \frac{n-1}{n} (1/2) \sigma_a^2$$

Empregando-se as variâncias derivadas, os coeficientes b_i são dados por:

$$b_1 = \frac{(1/4) \sigma_a^2 + (1/2) \sigma_a^2 / (nbm)}{\sigma_f^2 + \sigma_{fm}^2 / m + \sigma_e^2 / (mb) + \sigma_\delta^2 / (mnb)}$$

$$b_2 = \frac{(1/4) \sigma_a^2 + (1/2) \sigma_a^2 / (nbf)}{\sigma_m^2 + \sigma_{fm}^2 / f + \sigma_e^2 / (fb) + \sigma_\delta^2 / (fnb)}$$

$$b_3 = \frac{(1/2) \sigma_a^2 / (nb)}{\sigma_{fm}^2 + \sigma_e^2 / b + \sigma_\delta^2 / (nb)}$$

$$b_4 = \frac{(1/2) \sigma_a^2 / n}{\sigma_e^2 + \sigma_\delta^2 / n}$$

$$b_5 = \frac{(1/2) \sigma_a^2}{\sigma_\delta^2}$$

Os coeficientes b_i podem ser expressos em função apenas dos parâmetros h_a^2 , h_g^2 e c^2 :

$$b_1 = \frac{\left(\frac{nbm+2}{4} \right) h_a^2}{1 + (n-1) t_1 + (b-1) nt_2 + (m-2) nbt_3}$$

$$b_2 = \frac{\left(\frac{nbf+2}{4} \right) h_a^2}{1 + (n-1) t_1 + (b-1) nt_2 + (f-2) nbt_3}$$

$$b_3 = \frac{(1/2) h_a^2}{1 + (n-1) t_1 + (b-1) nt_2 - 2nbt_3}$$

$$b_4 = \frac{(1/2) h_a^2}{1 + (n-1) t_1 - nt_2}$$

$$b_5 = \frac{(1/2) h_a^2}{(1-t_1)}$$

Os efeitos aditivos dos genitores femininos podem ser preditos empregando-se a seguinte herdabilidade:

$$h_p^2 = \frac{\left(\frac{nbm}{2}\right) h_a^2}{1 + (n-1)t_1 + (b-1)nt_2 + (m-2)nbt_3}$$

Os efeitos aditivos dos genitores masculinos podem ser preditos por meio da mesma expressão de h_p^2 , porém substituindo m por f .

7.7.3. Índice multiefeitos para os efeitos de dominância

Nesta situação, o índice multiefeitos para predição dos efeitos de dominância é dado por:

$$I_2 = b_6 f_i + b_7 m_j + b_8 fm_{ij} + b_9 e_{ijk} + b_{10} \delta_{ijk\ell}.$$

Procedendo de forma análoga à realizada no tópico 7.7.2., obtém-se os b_i :

$$b_6 = \frac{(1/4) \sigma_d^2 / m + (3/4) \sigma_d^2 / (nbm)}{V_1}$$

$$b_7 = \frac{(1/4) \sigma_d^2 / f + (3/4) \sigma_d^2 / (nbf)}{V_2}$$

$$b_8 = \frac{\{[1 + (nb-1)(1/4)]/(nb)\} \sigma_d^2}{V_3}$$

$$b_9 = \frac{(3/4) \sigma_d^2 / n}{V_4}$$

$$b_{10} = \frac{(3/4) \sigma_d^2}{V_5}$$

Estes coeficientes b_i podem ser expressos em função apenas dos parâmetros h_a^2 , h_g^2 e c^2 :

$$b_6 = \frac{\left(\frac{nb+3}{4}\right) (h_g^2 - h_a^2)}{1 + (n-1)t_1 + (b-1)nt_2 + (m-2)nbt_3}$$

$$b_7 = \frac{\left(\frac{nb+3}{4}\right) (h_g^2 - h_a^2)}{1 + (n-1)t_1 + (b-1)nt_2 + (f-2)nbt_3}$$

$$b_8 = \frac{[1+(nb-1)(1/4)] (h_g^2 - h_a^2)}{1+(n-1) t_1 + (b-1) nt_2 - 2nbt_3}$$

$$b_9 = \frac{(3/4) (h_g^2 - h_a^2)}{1+(n-1) t_1 - nt_2}$$

$$b_{10} = \frac{(3/4) (h_g^2 - h_a^2)}{(1-t_1)}$$

7.7.4. BLUP para os efeitos aditivos e de dominância

Para o caso de delineamentos desbalanceados, deve-se empregar o procedimento BLUP para a predição dos efeitos aditivos e de dominância. No contexto dos modelos lineares mistos tem-se:

$y = X\beta + Za + Zd + Wc + e$, em que:

y, β, a, d, c e e : vetores de dados, de efeitos fixos (blocos), de efeitos genéticos aditivos (aleatório), de efeitos de dominância (aleatório), de efeitos de parcela (aleatório) e erros aleatórios, respectivamente.

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

As equações de modelo misto para a predição dos efeitos aditivos e de dominância pelo procedimento BLUP sob modelo individual (que devem fornecer os mesmos resultados que os índices I_1 e I_2 em situações de dados balanceados, já que é a sua versão matricial) equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{d} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\alpha_1 & Z'Z & Z'W \\ Z'X & Z'Z & Z'Z + D^{-1}\alpha_2 & Z'W \\ W'X & W'Z & W'Z & W'W + I\alpha_3 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ Z'y \\ W'y \end{bmatrix} \quad (1), \text{ em que:}$$

$$\alpha_1 = \frac{\sigma_{ad}^2}{\sigma_a^2} = \frac{1-h_g^2-c^2}{h_a^2}; \quad \alpha_2 = \frac{\sigma_{ad}^2}{\sigma_d^2} = \frac{1-h_g^2-c^2}{h_g^2-h_a^2}; \quad \alpha_3 = \frac{\sigma_{ad}^2}{\sigma_e^2} = \frac{1-h_g^2-c^2}{c^2}$$

A e D : matrizes de parentesco genético aditivo e de dominância, respectivamente.

$\sigma_{ad}^2 = (1-h_g^2-c^2) \sigma_y^2$: variância ambiental dentro da parcela; $\sigma_a^2 = h_a^2 \sigma_y^2$;

$\sigma_e^2 = c^2 \sigma_y^2$; $\sigma_d^2 = (h_g^2 - h_a^2) \sigma_y^2$.

Assim: $\alpha_1 = \frac{1-h_g^2-c^2}{h_a^2}$; $\alpha_2 = \frac{1-h_g^2-c^2}{h_g^2-h_a^2}$; $\alpha_3 = \frac{1-h_g^2-c^2}{c^2}$.

7.7.5. Índice multiefeitos e BLUP para efeitos genotípicos

Nesta situação, o índice multiefeitos para a predição dos efeitos genotípicos é dado por:

$$I_3 = (b_1 + b_6) f_i + (b_2 + b_7) m_j + (b_3 + b_8) fm_{ij} + (b_4 + b_9) e_{ijk} + (b_5 + b_{10}) \delta_{ijkl}.$$

Pelo procedimento BLUP, os valores genotípicos, dados por $\hat{g} = \hat{a} + \hat{d}$, podem ser preditos diretamente pelas equações de modelo misto:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + G^{-1}\sigma_{ad}^2 & Z'W \\ W'X & W'Z & W'W + I\alpha_3 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix} \quad (2), \text{ em que:}$$

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

7.7.6. Componentes de variância e dados utilizados

Para aplicação das metodologias, os seguintes parâmetros genéticos foram empregados:

$$h_a^2 = 0,25; \quad h_g^2 = 0,30; \quad c^2 = 0,08$$

Esses parâmetros estão associados aos seguintes coeficientes de correlação intraclasse e coeficientes α :

$$t_1 = 0,2175; \quad t_2 = 0,1375; \quad t_3 = 0,0625; \quad \alpha_1 = 2,48; \quad \alpha_2 = 12,40; \quad \alpha_3 = 7,75$$

Considerou-se, ainda, o seguinte subconjunto de dados, da espécie *Eucalyptus grandis*:

Indivíduo	Bloco	Macho (M)	Fêmea (F)	Progênie	Árvore	Parcela	Diâmetro (cm)
1	1	1	3	1-3	1	1	9,87
2	1	1	3	1-3	2	1	14,48
3	1	1	4	1-4	1	2	8,91
4	1	1	4	1-4	2	2	14,64
5	1	2	3	2-3	1	3	9,55
6	1	2	3	2-3	2	3	7,96
7	1	2	4	2-4	1	4	16,07
8	1	2	4	2-4	2	4	14,01
9	2	1	3	1-3	1	5	7,96
10	2	1	3	1-3	2	5	21,17
11	2	1	4	1-4	1	6	10,19
12	2	1	4	1-4	2	6	9,23
13	2	2	3	2-3	1	7	15,00
14	2	2	3	2-3	2	7	16,07
15	2	2	4	2-4	1	8	14,01
16	2	2	4	2-4	2	8	14,50

Os resultados da aplicação prática são apresentados a seguir.

Seleção por valores genéticos aditivos (propagação sexuada)

Os resultados referentes ao índice I_1 e BLUP (1) para os efeitos aditivos são apresentados na Tabela 7.17.

Tabela 7.17. Coeficientes de ponderação (b_i) das diferentes informações fenotípicas e efeitos aditivos preditos pelos procedimentos índice multiefeitos (I_1) e BLUP (1).

Ponderadores do Índice I_1	Indivíduos	Efeitos aditivos Preditos por I_1 e BLUP (1)
$b_1 = 0,41876$	1	-0,5279
$b_2 = 0,41876$	2	0,2085
$b_3 = 0,12594$	3	-0,6711
$b_4 = 0,13263$	4	0,2442
$b_5 = 0,15974$	5	-0,0857
	6	-0,3397
	7	0,7504
	8	0,4213
	9	-1,1074
	10	1,0029
	11	-0,6202
	12	-0,7736
	13	0,3915
	14	0,5624
	15	0,2330
	16	0,3113
	17*	-0,4489
	18**	0,4489
	19***	0,0209
	20****	-0,0209

* , ** , *** e **** : genitores 1, 2, 3 e 4, respectivamente.

Verifica-se, para o caso balanceado, a equivalência entre o BLUP e o índice multiefeitos. Adicionalmente, o BLUP fornece os valores genéticos aditivos preditos dos genitores (Tabela 7.17).

Os resultados referentes ao índice I_2 e BLUP (1) para os efeitos de dominância são apresentados na Tabela 7.18.

Tabela 7.18. Coeficientes de ponderação (b_i) das diferentes informações fenotípicas e efeitos de dominância preditos pelos procedimentos índice multiefeitos (I_2) e BLUP (1).

Ponderadores do Índice I_2	Indivíduos	Efeitos de dominância Preditos por I_2 e BLUP (1)
$b_6 = 0,05863$	1	-0,0510
$b_7 = 0,05863$	2	0,1700
$b_8 = 0,08816$	3	-0,2190
$b_9 = 0,03979$	4	0,0556
$b_{10} = 0,04792$	5	-0,1373
	6	-0,2135
	7	0,2469
	8	0,1482
	9	-0,2248
	10	0,4083
	11	-0,2037
	12	-0,2497
	13	0,0059
	14	0,0571
	15	0,0917
	16	0,1152
	17*	0,00
	18**	0,00
	19***	0,00
	20****	0,00

* , ** , *** e **** - parentais 1, 2, 3 e 4, respectivamente.

Os efeitos de dominância preditos para os genitores apresentaram valores zero, em função de os parentais não apresentarem observações próprias (Tabela 7.18). Adicionalmente à predição dos efeitos aditivos e de dominância de todos os indivíduos, o procedimento BLUP (1) forneceu, também, as estimativas das médias de bloco: $\hat{\beta}_1 = 11,9362$ e $\hat{\beta}_2 = 13,5162$ e as predições dos efeitos de parcela: $\hat{p}_1 = 0,0695$, $\hat{p}_2 = 0,0275$, $\hat{p}_3 = -0,5730$, $\hat{p}_4 = 0,4760$.

$$\hat{p}_5 = 0,2070 \quad \hat{p}_6 = -0,5913 \quad \hat{p}_7 = 0,3098 \quad \hat{p}_8 = 0,0745$$

Na Tabela 7.19 são apresentados os resultados referentes ao índice I_3 e procedimento BLUP (2), que fornecem as predições para os efeitos genotípicos.

Tabela 7.19. Coeficientes de ponderação das diferentes informações fenotípicas e efeitos genotípicos preditos pelos procedimentos índice multifeitos (I_3) e BLUP (2).

Ponderadores do Índice I_3	Indivíduos	Efeitos genotípicos
$b_1 + b_6 = 0,47739$	1	-0,5788
$b_2 + b_7 = 0,47739$	2	0,3785
$b_3 + b_8 = 0,21410$	3	-0,8901
$b_4 + b_9 = 0,17242$	4	0,2998
$b_5 + b_{10} = 0,20766$	5	-0,2230
	6	-0,5532
	7	0,9973
	8	0,5695
	9	-1,3322
	10	1,4111
	11	-0,8239
	12	-1,0233
	13	0,3974
	14	0,6196
	15	0,3248
	16	0,4265
	17*	-0,4489
	18**	0,4489
	19**	0,0209
	20****	-0,0209

*; **; *** e **** - parentais 1, 2, 3 e 4, respectivamente.

Para o caso de fatoriais desconexos e de pequena dimensão, deve-se incluir, nos índices multifeitos, o efeito de grupo de genitores (ver Capítulo 15) e no BLUP o efeito de grupo genético, conforme Resende & Rosa-Perez (1999a).

No caso de cruzamentos dialélicos, as equações de modelo misto associadas ao BLUP são idênticas às apresentadas para os cruzamentos fatoriais.

Considerações Finais

(i) Os procedimentos apresentados permitem a eficiente predição de efeitos aditivos, de dominância e genotípicos, associados a cruzamentos fatoriais e dialélicos; (ii) para o pleno uso dos índices e equações de modelo misto apresentados, são necessárias estimativas dos parâmetros herdabilidade no sentido restrito, herdabilidade no sentido amplo e correlação devida ao ambiente comum da parcela.

7.8. BLUP Individual e Índice Multifeitos Associados ao Látice

A eficiência do melhoramento genético de caracteres quantitativos depende da utilização de eficientes delineamentos de cruzamento e delineamentos experimentais, além da posterior utilização de procedimentos de seleção que

explorem todos os atributos favoráveis destes delineamentos. Neste contexto, o procedimento de seleção BLUP (melhor predição linear não viciada) (Henderson, 1984) desempenha papel relevante conduzindo a previsões acuradas e não viciadas de valores genéticos.

Quanto aos delineamentos experimentais, os mais utilizados e recomendados para a experimentação no campo são os de blocos casualizados e os de blocos incompletos, sendo os de blocos incompletos especialmente indicados na situação de grande número de tratamentos e alta variabilidade ambiental na área experimental. No melhoramento florestal, o delineamento de blocos casualizados tem sido o mais utilizado na América do Norte (Fu et al., 1998), os blocos incompletos têm sido os mais empregados na Austrália, na África do Sul e na Ásia (Williams & Matheson, 1994), e, no Brasil, ambos os tipos de delineamentos têm sido adotados.

A eficiência relativa entre os delineamentos experimentais depende, sobretudo, do nível de variação ambiental espacial na área experimental. Empregando um modelo geoestatístico espacial, o qual permite a especificação de vários níveis de variação ambiental, Fu et al. (1998) concluíram que os delineamentos de blocos incompletos (látice e alfa) são superiores em um grande número de situações, em termos de eficiência estatística para estimativa de médias de tratamentos. A seguir são apresentados os desenvolvimentos realizados por Resende (1999b) e Resende & Fernandes (2000).

7.8.1. Delineamento em látice e obtenção de médias ajustadas

Os delineamentos em blocos incompletos foram desenvolvidos por Yates (1936) e caracterizam-se por apresentar estratos experimentais (blocos) dentro de repetições. Dentre os delineamentos de blocos incompletos, os mais empregados na experimentação agropecuária são os látices, nos quais cada tratamento aparece uma única vez em cada repetição e cada par de tratamentos, no máximo, uma vez em cada bloco. Os látices podem ser balanceados (quando cada tratamento ocorre uma vez com qualquer outro tratamento no mesmo bloco), ou parcialmente balanceados (quando apenas alguns tratamentos ocorrem juntos no mesmo bloco). Nos látices balanceados, os ajustes das médias de tratamento são igualmente eficientes para todos os tratamentos, fato que não ocorre no látice parcialmente balanceado (Miranda Filho, 1987).

No presente tópico será considerado o látice quadrado (k^2 tratamentos e k tratamentos por bloco) balanceado ($k+1$ repetições), por ser o tipo mais utilizado na avaliação de um grande número de progênies em muitas repetições, conforme prática do melhoramento florestal. Os aspectos teóricos e práticos fundamentais associados aos delineamentos em látice são apresentados nas obras de Kempthorne (1952), Federer (1955), Malheiros (1982), Iemma (1987) e Hinckelman & Kempthorne (1994).

Análise intrablocos

O modelo estatístico ao nível de médias de parcela equivale a:

$$Y_{ijq} = \mu + r_i + b_{ij} + p_q + e_{ijq}, \text{ em que:}$$

μ , r_i , b_{ij} e p_q : são os efeitos da média, repetição, bloco dentro de repetições e progênies (ou tratamentos), respectivamente, assumidos como fixos (Kempthorne, 1952) e e_{ijq} é o erro intrabloco, aleatório, com distribuição normal, independentemente distribuído com média zero e variância σ^2 .

As estimativas de parâmetros são obtidas pela minimização de $\sum (Y_{ijq} - \mu - r_i - b_{ij} - p_q)^2$, sujeita às restrições $\sum_i r_i = 0$; $\sum_j b_{ij} = 0$ (para todo i); $\sum_q p_q = 0$.

Em associação com este modelo, as composições de médias são:

– Média do tratamento q

$$\bar{Y}_{..q} = \mu + \frac{\sum b_{ij}}{(k+1)} + p_q, \text{ em que } \sum b_{ij} \text{ refere-se aos blocos que contêm o tratamento } q.$$

– Média do bloco j dentro da repetição i

$$\bar{Y}_{ij} = \mu + r_i + b_{ij} + \frac{\sum p_q}{k}, \text{ em que } \sum p_q \text{ refere-se aos tratamentos que ocorrem no bloco } ij.$$

– Média de todos os blocos que contêm o tratamento q

$$\bar{Y}_{...}^* = \mu + \frac{\sum b_{ij}}{(k+1)} + \frac{p_q}{(k+1)} \quad (\text{visto que } \sum_q p_q = 0 \text{ e que os demais tratamentos ocorrem apenas uma vez cada um, em todos os blocos em que se verifica o tratamento } q).$$

Com base nestes componentes de médias, o efeito de tratamento ajustado para os efeitos de blocos, considerando uma análise intrablocos (Yates, 1936), é dado por:

$$\bar{Y}_{..q} - \bar{Y}_{...}^* = \mu + \frac{\sum b_{ij}}{(k+1)} + p_q - \mu - \frac{\sum b_{ij}}{(k+1)} - \frac{p_q}{(k+1)}$$

$$\bar{Y}_{..q} - \bar{Y}_{...}^* = p_q - \frac{p_q}{(k+1)}$$

$$\frac{(k+1)}{k} (\bar{Y}_{..q} - \bar{Y}_{...}^*) = p_q$$

Assim, $p_q = \frac{(k+1)}{k} (\bar{Y}_{..q} - \bar{Y}_{...}^*)$ fornece o efeito de tratamento, livre do confundimento devido aos efeitos de blocos. A média de tratamentos ajustada, por sua vez, é dada por $\bar{Y}_{..q}^* = \bar{Y}_{...} + \frac{k+1}{k} (\bar{Y}_{..q} - \bar{Y}_{...}^*)$, em que $\bar{Y}_{...}$ é a média geral do experimento.

Verifica-se, portanto, que a análise intrablocos é um procedimento estatístico exato e baseia-se no procedimento de quadrados mínimos ordinários.

Análise com recuperação da informação interblocos

Na análise intrabloco somente, as comparações entre parcelas do mesmo bloco são usadas nas estimativas dos efeitos de tratamentos. Entretanto, um outro tipo de análise, baseada na recuperação da informação interblocos (Yates, 1940), aproveita também as comparações entre blocos na estimação dos efeitos de tratamentos (Pimentel Gomes, 1987).

Na análise com recuperação da informação interblocos, os efeitos de blocos dentro de repetição são considerados aleatórios, normalmente e independentemente distribuídos com média zero, variância σ_b^2 e independentes de e_{ijq} . Neste caso, o ajuste das médias combina as informações dos contrastes de parcela dentro de blocos e dos contrastes entre blocos. O procedimento, porém, não é estatisticamente exato mas apenas aproximado devido à suposição de σ^2 e σ_b^2 conhecidos e que na prática devem ser estimados (Cochran & Cox, 1957; Malheiros & Godoi, 1984a e b). Denominando $w=1/\sigma^2$ e $w'=1/(\sigma^2+k\sigma_b^2)$, os pesos dados às fontes de informação de parcela dentro de bloco e entre blocos (w'/k), respectivamente, as médias ajustadas são dadas por:

$$\bar{Y}_{..q}^{**} = \frac{Y_{..q}}{(k+1)} + \frac{(w-w') [k Y_{..q} - (k+1) \bar{Y}_{...}^* + \bar{Y}_{...}]}{k (k+1) (kw+w')}$$

$$\bar{Y}_{..q}^{**} = \bar{Y}_{..q} + (w-w') \frac{\bar{Y}_{..q} - (k+1) \bar{Y}_{...}^* + K \bar{Y}_{...}}{(kw+w')}$$

Verifica-se então que $\bar{Y}_{..q}^*$ e $\bar{Y}_{..q}^{**}$ são iguais (isto é, os ajustes são iguais) se $w'=0$, ou seja, quando σ_b^2 for muito alto. Nos demais casos, com a análise interblocos o ajustamento é sempre menor. Por outro lado, o ajustamento é nulo quando $w'=w$, ou seja quando $\sigma_b^2=0$. Neste último caso, não existem diferen-

ças entre blocos dentro de repetições e a análise reduz-se à do delineamento de blocos completos casualizados.

Sendo $(\sigma^2 + \sigma_b^2)$ a esperança da variância do erro associada ao delineamento de blocos casualizados e σ^2 a esperança da variância do erro intrabloco, a eficiência do látice sobre o delineamento de blocos ao acaso pode ser dada por $(\sigma^2 + \sigma_b^2)/\sigma^2$. A alternativa é a expressão desta eficiência em termos do erro efetivo (para teste de médias de tratamentos ajustados com base na análise com recuperação da informação interblocos), onde tem-se que a eficiência é dada por $(\sigma^2 + \sigma_b^2)/[\sigma^2(1 + \sigma_b^2/(\sigma^2 + k\sigma_b^2))]$.

7.8.2. BLUP e índice multiefeitos ajustando bloco dentro de repetição como efeito fixo

O modelo estatístico ao nível de indivíduo é:

$y = X\beta + Za + Tc + \delta$, em que:

y , β , a , c e δ : vetores de dados, de efeitos fixos (médias de blocos dentro de repetições), de valores genéticos aditivos (aleatório), de efeitos de parcela (aleatório) e erros aleatórios, respectivamente.

X , Z e T : matrizes de incidência para β , a e c , respectivamente.

As equações de modelo misto para predição de valores genéticos pelo procedimento BLUP sob modelo individual equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'T \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'T \\ T'X & TZ & T'T + I\lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ T'y \end{bmatrix} \quad (1), \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_{rd}^2}{\sigma_a^2} = \frac{1 - h_a^2 - c^2}{h_a^2}; \quad \lambda_2 = \frac{\sigma_{rd}^2}{\sigma_e^2} = \frac{1 - h_a^2 - c^2}{c^2}$$

$\sigma_{rd}^2 = \sigma_\delta^2 - (1 - \rho_a)$ σ_a^2 : variância residual (ambiental + não aditiva) dentro de parcela.

σ_e^2 : variância entre parcelas.

σ_a^2 : variância genética aditiva.

$h_a^2 = \frac{\sigma_a^2}{\sigma_y^2}$: herdabilidade individual no sentido restrito.

$c^2 = \frac{\sigma_e^2}{\sigma_y^2}$: fração da correlação intraclasse entre indivíduos de uma mesma parcela, devida ao ambiente comum da parcela.

$\sigma_y^2 = \sigma_\delta^2 + \sigma_e^2 + \sigma_p^2 = \sigma_{rd}^2 + \sigma_e^2 + \sigma_a^2$: variância fenotípica.

σ_δ^2 : variância entre plantas dentro de parcela.

σ_p^2 : variância entre progênies.

A : matriz de parentesco genético aditivo entre os indivíduos.

Para o caso de dados balanceados (sobrevivência 100%), as soluções para \hat{A} são equivalentes às fornecidas pelo seguinte índice multiefeitos (que considera todos os efeitos do modelo estatístico, exceto os efeitos da média geral, de repetição e de bloco dentro de repetição):

$$I_1 = b_1 \delta_{ijq\ell} + b_2 e_{ijq} + b_3 p_q$$

$= b_1(Y_{ijq\ell} - \bar{Y}_{ijq}) + b_2(\bar{Y}_{ijq} - \bar{Y}_{..q} - \bar{Y}_{ij..}^B + \bar{Y}_{...}^{*B}) + b_3(\bar{Y}_{..q} - \bar{Y}_{...}^{*B})$, em que:

$$b_1 = h_\delta^2 = \frac{(1-\rho_a) \sigma_a^2}{\sigma_\delta^2} = \frac{(1-\rho_a) h_a^2}{(1-t_1)}$$

$$b_2 = h_p^2 = \frac{[(1-\rho_a)/n] \sigma_a^2}{\sigma_e^2 + \sigma_\delta^2/n} = \frac{(1-\rho_a) h_a^2}{1+(n-1) t_1 - nt_2}$$

$$b_3 = h_m^2 = \frac{\{[1+(nk+n-1) \rho_a]/nb\} \sigma_a^2}{\sigma_p^2 + \sigma_e^2/(k+1) + \sigma_\delta^2/(nk+n)} = \frac{[1+(nk+n-1) \rho_a] h_a^2}{1+(n-1) t_1 + knt_2}$$

$t_2 = \rho_a h_a^2$: correlação fenotípica intraclasse entre os indivíduos da progénie, em diferentes parcelas.

$t_1 = t_2 + c^2$: correlação fenotípica intraclasse entre os indivíduos de uma mesma parcela.

h_δ^2 , h_p^2 e h_m^2 : herdabilidades dos efeitos de indivíduos dentro de parcela, de progênies e de parcelas, respectivamente.

ρ_a : correlação genética intraclasse entre os indivíduos da progénie ($\rho_a = 0,25$ para progênies de meios-irmãos e $\rho_a = 0,5$ para progênies de irmãos gêmeos).

n e k : número de plantas por parcela e de blocos dentro de repetição, respectivamente.

$Y_{ijq\ell}$, \bar{Y}_{ijq} , $\bar{Y}_{..q}$: valor individual e médias de parcela e de progênies, respectivamente.

$\bar{Y}_{ij..}^B$: média do bloco j dentro da repetição i obtida pelo procedimento de quadrados mínimos generalizados (ou do próprio vetor β), a qual é uma melhor estimativa linear não viciada (BLUE).

$\bar{Y}_{...}^{*B}$: média de todos os blocos que contêm o tratamento q , obtida pelo procedimento de quadrados mínimos generalizados, a qual é também BLUE.

As médias de tratamentos (progénies) ajustadas, obtidas pelo procedimento BLUP são dadas por:

$$\bar{Y}_{..q.}^{***} = \hat{a}_p / h_{pp}^2 + \bar{Y}_{...} , \text{ em que:}$$

\hat{a}_p : valor genético predito do genitor, obtido por $\hat{a}_p = h_{pp}^2 (\bar{Y}_{..q.} - \bar{Y}_{...}^{*B})$,

$\bar{Y}_{...}$: média geral do experimento,

$$h_{pp}^2 = \frac{(1/2) \sigma_a^2}{\sigma_p^2 + \sigma_e^2 / (k+1) + \sigma_\delta^2 / (nk+n)} : \text{herdabilidade para a seleção de genitores.}$$

7.8.3. BLUP ajustando bloco dentro de repetições como efeito aleatório

Neste caso, os efeitos b_{ij} são considerados aleatórios e o modelo linear misto equivale a:

$$y = X^* \beta^* + Za + Tc + Sg + \delta, \text{ em que:}$$

β^* e g : vetores de efeitos fixos (médias de repetições) e dos efeitos de bloco (aleatório), respectivamente.

X^* e S : matrizes de incidência para β^* e g , respectivamente.

As equações de modelo misto equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{c} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X^{*'} X^* & X^{*'} Z & X^{*'} T & X^{*'} S \\ Z' X^* & Z' Z + A^{-1} \lambda_1 & Z' T & Z' S \\ T' X^* & T' Z & T' T + I \lambda_2 & T' S \\ S' X^* & S' Z & S' T & S' S + I \lambda_3 \end{bmatrix}^{-1} \begin{bmatrix} X^{*'} y \\ Z' y \\ T' y \\ S' y \end{bmatrix} \quad (2), \text{ em que:}$$

$$\lambda_3 = \frac{\sigma_{rd}^2}{\sigma_b^2} = \frac{1 - h_a^{2*} - c^{2*} - b^2}{b^2}$$

$b^2 = \frac{\sigma_b^2}{\sigma_{y^*}^2}$: correlação intraclasse entre indivíduos de um mesmo bloco, mas de diferentes parcelas.

$$h_a^2* = \frac{\sigma_a^2}{\sigma_{y^*}^2} : \text{herdabilidade no sentido restrito.}$$

$$c^2* = \frac{\sigma_e^2}{\sigma_{y^*}^2} : \text{correlação intraclasse devido ao ambiente comum da parcela.}$$

$$\sigma_{y^*}^2 = \sigma_p^2 + \sigma_e^2 + \sigma_\delta^2 + \sigma_b^2 : \text{variância fenotípica individual.}$$

Para o caso de dados balanceados (sobrevivência 100%), as soluções para \hat{a} são eqüivalentes às fornecidas pelo seguinte índice multiefeitos:

$$I_2 = b_1(Y_{ijq\ell} - \bar{Y}_{ijq.}) + b_2(\bar{Y}_{ijq.} - \bar{Y}_{..q.} - \bar{Y}_{ij..} + \bar{Y}^*) + b_3(\bar{Y}_{..q.} - \bar{Y}_{...}) + \\ + b_4(\bar{Y}_{ij..} - \bar{Y}_{i...}) + b_5(\bar{Y}^* - \bar{Y}_{...})$$

$\bar{Y}_{ij..}$, $\bar{Y}_{i...}$ e \bar{Y}^* : média do bloco j dentro da repetição i , média da repetição i e média de todos os blocos que contêm o tratamento q , respectivamente, obtidas pelo método de quadrados mínimos ordinários.

Como existem covariâncias entre o quarto e o quinto componentes do índice I_2 , os pesos b_4 e b_5 são determinados por:

$$\begin{bmatrix} b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} \sigma_{c_4} & \sigma_{c_4 c_5} \\ \sigma_{c_5 c_4} & \sigma_{c_5} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{c_4 a} \\ \sigma_{c_5 a} \end{bmatrix}, \text{ em que}$$

$$\sigma_{c_4}^2 = Var(\bar{Y}_{ij..} - \bar{Y}_{i...}) = \frac{k-1}{k} \left(\sigma_b^2 + \frac{\sigma_p^2}{k} + \frac{\sigma_e^2}{k} + \frac{\sigma_\delta^2}{nk} \right)$$

$$\sigma_{c_5}^2 = Var(\bar{Y}^* - \bar{Y}_{...}) = \frac{k-1}{k} \left(\frac{\sigma_b^2}{k+1} + \frac{\sigma_p^2}{k} + \frac{\sigma_e^2}{k(k+1)} + \frac{\sigma_\delta^2}{nk(k+1)} \right)$$

$$\sigma_{c_4 c_5} = \sigma_{c_5 c_4} = \sigma_{c_5}^2$$

$$\sigma_{c_4 a}^2 = \frac{k-1}{k} \frac{1+(n-1) \rho_a}{nk} \sigma_a^2$$

$$\sigma_{c_5 a}^2 = \frac{k-1}{k} \frac{1+(nk+n-1) \rho_a}{nk(k+1)}$$

k : número de blocos dentro de repetição.

As médias de tratamentos ajustadas, obtidas pelo procedimento BLUP são dadas por:

$\bar{Y}_{..q}^{***} = \hat{a}_p / h_{pp}^2 + \bar{Y}_{...}$, em que \hat{a}_p é obtido por:

$\hat{a}_p = h_{pp}^2 (\bar{Y}_{..q} - \bar{Y}_{...}) + h_{bt}^2 (\bar{Y}_{...}^* - \bar{Y}_{...})$, em que:

$$h_{bt}^2 = \frac{\frac{(k-1)}{k} \frac{0,5 \sigma_a^2}{k}}{\frac{k-1}{k} \left(\frac{\sigma_b^2}{k+1} + \frac{\sigma_p^2}{k} + \frac{\sigma_e^2}{k(k+1)} + \frac{\sigma_\delta^2}{nk(k+1)} \right)}.$$

7.8.4. Acurácia e variância do erro de predição dos valores genéticos

Sendo a matriz dos coeficientes das equações de modelo misto:

$$M_1 = \begin{bmatrix} X'X & X'Z & X'T \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'T \\ T'X & T'Z & T'T + I\lambda_2 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \text{ para o modelo (1) e}$$

$$M_2 = \begin{bmatrix} X^{*'}X^* & X^{*'}Z & X^{*'}T & X^{*'}S \\ Z^{*'}X^* & Z^{*'}Z + A^{-1}\lambda_1 & Z^{*'}T & Z^{*'}S \\ T^{*'}X^* & T^{*'}Z & T^{*'}T + I\lambda_2 & T^{*'}S \\ S^{*'}X^* & S^{*'}Z & S^{*'}T & S^{*'}S + I\lambda_3 \end{bmatrix} = \begin{bmatrix} M_{11}^* & M_{12}^* & M_{13}^* & M_{14}^* \\ M_{21}^* & M_{22}^* & M_{23}^* & M_{24}^* \\ M_{31}^* & M_{32}^* & M_{33}^* & M_{34}^* \\ M_{41}^* & M_{42}^* & M_{43}^* & M_{44}^* \end{bmatrix}$$

para o modelo (2), têm-se as inversas

$$M_1^{-1} = \begin{bmatrix} M^{11} & M^{12} & M^{13} \\ M^{21} & M^{22} & M^{23} \\ M^{31} & M^{32} & M^{33} \end{bmatrix} \text{ e}$$

$$M_2^{-1} = \begin{bmatrix} M^{*11} & M^{*12} & M^{*13} & M^{14} \\ M^{*21} & M^{*22} & M^{*23} & M^{24} \\ M^{*31} & M^{*32} & M^{*33} & M^{34} \\ M^{*41} & M^{*42} & M^{*43} & M^{44} \end{bmatrix}$$

As variâncias dos erros de predição (PEV) (Henderson, 1975) dos valores genéticos, quando R (a matriz dos resíduos) = $I\sigma_{rd}^2$, são dadas por

$$PEV_{(1)} = \text{Var}(\hat{a} - a) = M^{22} \sigma_{rd}^2 \quad \text{para o modelo (1) e}$$

$$PEV_{(2)} = \text{Var}(\hat{a} - a) = M^{*22} \sigma_{rd}^2 \text{ para o modelo (2).}$$

As acurárias (correlação entre valores genéticos preditos e verdadeiros) associadas à predição dos valores genéticos são dadas por $r_{\hat{a}a(1)} = [1 - PEV_{(1)} / \sigma_a^2]^{1/2}$ para o modelo (1) e $r_{\hat{a}a(2)} = [1 - PEV_{(2)} / \sigma_a^2]^{1/2}$ para o modelo (2).

Estimadores equivalentes são $r_{\hat{a}a} = (1 - d_i \lambda_i)^{1/2}$ e $PEV = (1 - r_{\hat{a}a}^2) \sigma_a^2$, em que d_i é o i -ésimo elemento da diagonal de M^{22} ou M^{*22} .

7.8.5. Análise considerando o delineamento de blocos ao acaso

A predição de valores genéticos considerando o delineamento de blocos ao acaso pode ser feita pelo procedimento BLUP ou por meio do índice:

$$I_3 = b_1(Y_{ijq\ell} - \bar{Y}_{ijq.}) + b_6(\bar{Y}_{ijq.} - \bar{Y}_{..q.} + \bar{Y}_{...}) + b_7(\bar{Y}_{..q.} - \bar{Y}_{...}), \text{ em que:}$$

$$b_6 = \frac{(1 - \rho_a)}{\frac{n}{\sigma_e^2 + k\sigma_b^2 + \sigma_\delta^2/n}} \sigma_a^2; \quad b_7 = \frac{\frac{1 + (nk + n - 1)\rho_g}{n(k-1)}}{\sigma_p^2 + (\sigma_e^2 + K\sigma_b^2)/(k+1) + \sigma_\delta^2/[n(k+1)]} \sigma_a^2$$

7.8.6. Exemplo prático

Foram considerados, para exercício das metodologias, os seguintes componentes de variância:

$$\sigma_a^2 = 1,0888 \quad \sigma_e^2 = 0,5187; \quad \sigma_b^2 = 0,30; \quad \sigma_p^2 = 0,2722; \quad \sigma_\delta^2 = 5,8672 \quad \sigma_{nd}^2 = 5,0506$$

$$\sigma_y^2 = 6,6584; \quad \sigma_y^{*2} = 6,9584$$

$$h_a^2 = 0,1635; \quad h_a^{*2} = 0,15648$$

$$c^2 = 0,0779; \quad c^{*2} = 0,07455$$

$$b^2 = 0,043115$$

$$\alpha_1 = 4,6387; \quad \alpha_2 = 9,7370 \quad \alpha_3 = 16,8353$$

$$w = 1/(\sigma_\delta^2/n + \sigma_e^2) = 0,28966 \quad w' = 1/(\sigma_\delta^2/n + \sigma_e^2 + K\sigma_b^2) = 0,24677.$$

Utilizou-se, também, o seguinte subconjunto de dados associados a um teste de progénies de meios-irmãos ($\rho_a = 0,25$) de *Eucalyptus grandis*, o qual corresponde a um látice balanceado com $k^2 = 4$ tratamentos, $(k + 1) = 3$ repetições e $k = 2$ blocos por repetição.

Indivíduo	Repetição	Bloco	Tratamento (Progénie)	Indivíduos/ Parcela	Diâmetro (cm)
1	1	1	1	1	9,87
2	1	1	1	2	10,00
3	1	1	2	1	14,48
4	1	1	2	2	15,00
5	1	2	3	1	8,91
6	1	2	3	2	9,00
7	1	2	4	1	15,00
8	1	2	4	2	14,64
9	2	3	1	1	9,55
10	2	3	1	2	10,00
11	2	3	3	1	8,00
12	2	3	3	2	7,96
13	2	4	2	1	15,00
14	2	4	2	2	16,07
15	2	4	4	1	14,01
16	2	4	4	2	14,50
17	3	5	1	1	7,50
18	3	5	1	2	9,30
19	3	5	4	1	10,30
20	3	5	4	2	17,00
21	3	6	2	1	13,50
22	3	6	2	2	15,00
23	3	6	3	1	18,00
24	3	6	3	2	12,00

Os resultados da aplicação prática são apresentados na seqüência.

Médias de tratamentos ajustadas pelo látice e BLUP

A eficiência do látice em relação ao delineamento de blocos ao acaso foi equivalente a 1,087 (8,7%) para a análise intrablocos e 1,012 (1,2%) para a análise com recuperação da informação interblocos. Na Tabela 7.20, são apresentadas as médias ajustadas obtidas pelos diferentes procedimentos.

Tabela 7.20. Médias ajustadas obtidas pela análise em blocos ao acaso ($\bar{Y}_{..q}$), pelo látice mediante a análise intrablocos ($\bar{Y}_{..q}^*$), pelo látice por meio da análise com recuperação da informação interblocos ($\bar{Y}_{..q}^{**}$), pelo BLUP considerando bloco dentro de repetição, como efeito fixo ($\bar{Y}_{..q}^{***}$) e pelo BLUP considerando bloco dentro de repetição como efeito aleatório ($\bar{Y}_{..q}^{****}$), bem como proporções entre elas.

Tratamento	$\bar{Y}_{..q}$	$\bar{Y}_{..q}^*$	$\bar{Y}_{..q}^{**}$	$\bar{Y}_{..q}^{***}$	$\bar{Y}_{..q}^{****}$	$\bar{Y}_{..q}^*/\bar{Y}_{..q}^{***}$	$\bar{Y}_{..q}^{**}/\bar{Y}_{..q}^{****}$
1	9,37	10,21	9,46	10,80	9,57	0,9454	0,9885
2	14,84	13,61	14,71	13,22	14,62	1,0295	1,0062
3	10,65	10,55	10,63	11,04	10,70	0,9556	0,9935
4	14,24	14,73	14,29	14,03	14,21	1,0499	1,0056

Verificam-se valores de médias ajustadas mais próximas pelas análises em látice com recuperação da informação interblocos e BLUP ajustando bloco dentro de repetição como efeito aleatório, conforme esperado teoricamente. Os resultados foram praticamente coincidentes, conforme pode ser observado pelos valores da relação $\bar{Y}_{..q}^{**}/\bar{Y}_{..q}^{***}$ (Tabela 7.20).

Comparando-se os resultados do látice pela análise intrablocos com o BLUP ajustando bloco dentro de repetição como efeito fixo, verifica-se que os valores obtidos diferem, conforme depreende-se da relação $\bar{Y}_{..q}^*/\bar{Y}_{..q}^{***}$ (Tabela 7.20). Embora conceitualmente semelhantes por considerarem os efeitos de blocos como fixos, estes dois métodos diferem bastante na estimação destes efeitos visto que o BLUP estima-os por meio do método de quadrados mínimos generalizados - GLS - e o látice, pelo método de quadrados mínimos ordinários - OLS. Na estimação dos efeitos de blocos por GLS, o BLUP considera o nível genético dos indivíduos (e progénies) que se desenvolvem em cada bloco, fato que não é considerado pelo método OLS. Assim, neste caso, o procedimento BLUP é superior à análise do látice, pois utiliza estimativas não viciadas (pelo valor genético dos indivíduos) dos efeitos de blocos. As estimativas das médias de blocos utilizados pelos dois procedimentos são apresentadas na Tabela 7.21.

Tabela 7.21. Estimativas das médias de blocos utilizadas pelo procedimento látice ($\bar{Y}_{ij..}$) e BLUP ajustando bloco como efeito fixo ($\bar{Y}_{ij..}^B$) bem como a média de todos os blocos que contêm um tratamento, obtidos pelos procedimentos látice ($\bar{Y}_{...}^*$) e BLUP ($\bar{Y}_{...}^{*B}$).

Bloco	$\bar{Y}_{ij..}$	$\bar{Y}_{ij..}^B$	Tratamento	$\bar{Y}_{...}^*$	$\bar{Y}_{...}^{*B}$
1	12,3375	12,3873	1	10,7467	10,8404
2	11,8875	11,8377	2	13,9525	13,8919
3	8,8775	9,1358	3	11,7967	11,8751
4	14,8950	14,6367	4	12,6025	12,4909
5	11,0250	10,9982			
6	14,6250	14,6518			

Os valores das médias de bloco $\bar{Y}_{ij..}$ foram estimados pelo BLUP (1) por meio da expressão $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$, em que $V = ZAZ'$ $\sigma_a^2 + WI\sigma_e^2W' + I\sigma_{rd}^2 = C + WI\sigma_e^2W' + I\sigma_{rd}^2$. A matriz V apresenta os valores de σ_y^2 na diagonal, $[1/4(\sigma_a^2 + \sigma_e^2)]$ nas associações (fora da diagonal) entre indivíduos da mesma parcela e $(1/4)\sigma_a^2$ nas associações entre indivíduos meios-irmãos em diferentes parcelas. Verifica-se, assim, que a matriz V permite considerar as correlações entre os efeitos aleatórios, na estimação dos efeitos fixos.

O procedimento BLUP (2) propicia uma estimação BLUE dos efeitos de repetição, mas como cada repetição comporta todos os tratamentos (progênies), as estimações GLS e OLS destes efeitos são equivalentes neste caso, fato que conduziu aos resultados similares entre os procedimentos BLUP (2) e análise interblocos do látice.

Conclui-se, em princípio, que a análise intrablocos do látice não é adequada, pois, mesmo para um delineamento balanceado, conduz a médias ajustadas viciadas. Uma vez que a análise do látice com recuperação da informação interblocos e BLUP que considera o efeito de blocos como aleatório praticamente se equivalem, conclui-se que este tipo de análise do látice pode ser aplicado na formulação de índices multiefeitos para a seleção em caso de dados平衡ados, embora possam conduzir a resultados viciados. Entretanto, para a situação de grandes experimentos, outras considerações são necessárias, conforme discutido no tópico a seguir.

7.8.7. Modelos alternativos do BLUP associado ao látice

Os resultados referentes aos índices I_1 e I_2 e procedimentos BLUP (1) e BLUP (2), bem como seus ponderadores, são apresentados na Tabela 7.22.

Tabela 7.22. Coeficientes de ponderação (b e h^2) das diferentes informações fenotípicas e valores genéticos preditos por diferentes procedimentos.

Ponderadores do Índice I_1	Ponderadores do Índice I_2	Ponderadores associados ao BLUP (1) e BLUP (2)	Indivíduos	Valores genéticos preditos pelo Índice I_1 e BLUP (1)	Valores genéticos preditos pelo Índice I_2 e BLUP (2)
$b_1=0,1392$	$b_1=0,1392$	$h_{pp}^2=0,38260$	1	-0,5471	-0,7272
$b_2=0,1183$	$b_2=0,1183$	$h_{bt}^2=0,33544$	2	-0,5290	-0,7091
$b_3=0,2869$	$b_3=0,2869$		3	0,4022	0,6648
	$b_4=0,1008$		4	0,4746	0,7372
	$b_5=0,1508$		5	-0,5547	-0,6406
			6	-0,5421	-0,6280
			7	0,6731	0,6765
			8	0,6230	0,6264

Continua...

Tabela 7.22. Continuação.

Ponderadores do Índice I_1	Ponderadores do Índice I_2	Ponderadores associados ao BLUP (1) e BLUP (2)	Indivíduos	Valores genéticos preditos pelo Índice I_1 e BLUP (1)	Valores genéticos preditos pelo Índice I_2 e BLUP (2)
$b_1=0,1392$	$b_1=0,1392$	$h_{pp}^2 = 0,38260$	9	-0,2037	-0,6915
$b_2=0,1183$	$b_2=0,1183$	$h_{bt}^2 = 0,33544$	10	-0,1411	-0,6289
$b_3=0,2869$	$b_3=0,2869$		11	-0,3414	-0,6791
	$b_4=0,1008$		12	-0,3470	-0,6847
	$b_5=0,1508$		13	0,1920	0,7064
			14	0,3409	0,8553
			15	0,2161	0,5272
			16	0,2843	0,5954
			17	-0,6806	-1,0744
			18	-0,4300	-0,8239
			19	0,1427	-0,0118
			20	1,0752	0,9207
			21	0,0083	0,4287
			22	0,2170	0,6375
			23	0,2512	0,3791
			24	-0,5838	-0,4560
			25*	-0,5626	-1,0345
			26*	0,3633	0,8955
			27*	-0,4706	-0,6021
			28*	0,6698	0,7410

* referente às quatro genitoras.

Verifica-se que a consideração dos efeitos de blocos dentro de repetição como fixos ou aleatórios conduz a resultados bastante diferenciados (Tabela 7.22). Isto ocorre devido à estimativa não viciada (BLUE) dos efeitos de blocos quando estes são considerados fixos, fato que não ocorre quando tais efeitos são considerados aleatórios e, também, devido ao fato de que frações adicionais da variação genética aditiva são utilizadas na predição quando os efeitos de blocos são considerados aleatórios.

Na Tabela 7.23 são apresentadas as variâncias do erro de predição (PEV) dos valores genéticos, bem como as acurárias associadas aos dois modelos de predição (BLUP).

Tabela 7.23. Variâncias dos erros de predição associadas aos modelos BLUP (1) (PEV_1) e BLUP (2) (PEV_2), bem como acurárias associadas ao modelo BLUP (1) ($r_{\hat{aa}_1}$) e BLUP (2) ($r_{\hat{aa}_2}$).

Indivíduos	PEV_1	PEV_2	$r_{\hat{aa}_1}$	$r_{\hat{aa}_2}$
1	0,9539	0,9247	0,3521	0,3882
2	0,9539	0,9247	0,3521	0,3882
3	0,9539	0,9247	0,3521	0,3882
4	0,9539	0,9247	0,3521	0,3882
5	0,9539	0,9247	0,3521	0,3882
6	0,9539	0,9247	0,3521	0,3882
7	0,9539	0,9247	0,3521	0,3882
8	0,9539	0,9247	0,3521	0,3882
9	0,9539	0,9247	0,3521	0,3882
10	0,9539	0,9247	0,3521	0,3882
11	0,9539	0,9247	0,3521	0,3882
12	0,9539	0,9247	0,3521	0,3882
13	0,9539	0,9247	0,3521	0,3882
14	0,9539	0,9247	0,3521	0,3882
15	0,9539	0,9247	0,3521	0,3882
16	0,9539	0,9247	0,3521	0,3882
17	0,9539	0,9247	0,3521	0,3882
18	0,9539	0,9247	0,3521	0,3882
19	0,9539	0,9247	0,3521	0,3882
20	0,9539	0,9247	0,3521	0,3882
21	0,9539	0,9247	0,3521	0,3882
22	0,9539	0,9247	0,3521	0,3882
23	0,9539	0,9247	0,3521	0,3882
24	0,9539	0,9247	0,3521	0,3882
25*	0,9776	0,9389	0,3196	0,3711
26*	0,9776	0,9389	0,3196	0,3711
27*	0,9776	0,9389	0,3196	0,3711
28*	0,9776	0,9389	0,3196	0,3711

* referente às 4 genitoras.

Verifica-se que o procedimento BLUP (1) conduziu a maiores variâncias do erro de predição (e, conseqüentemente, menor precisão) e menores acurárias do que o BLUP (2) (Tabela 7.23), fato que indica, a princípio, uma superioridade do BLUP (2), o qual, entretanto, pode ser viciado.

A consideração dos efeitos de blocos como fixos propicia predições invariantes aos efeitos fixos, removendo os vícios nas comparações genéticas através dos blocos. Tal consideração é, sobretudo, importante quando existe uma alocação não aleatória dos materiais genéticos nos blocos (fato que pode ocorrer quando o número de tratamentos é pequeno) e, neste caso, o BLUP minimiza o vício causado por associações entre efeitos de blocos e o nível genético

dos materiais que neles se desenvolvem. Quando não existem associações entre efeitos de blocos e o nível genético dos materiais, não há vícios (PEV aproximadamente igual ao erro quadrático médio) mesmo quando se consideram como aleatórios os efeitos de blocos.

Por outro lado, considerando aleatórios os efeitos de blocos, maior quantidade de informações é utilizada, fato que conduz a menores variâncias do erro de predição e maior acurácia. Entretanto, quando o tamanho do bloco é relativamente grande, ocorre pequeno decréscimo na PEV, fato que pode ser visualizado pela expressão $\sigma_{c_{4a}} = \frac{1+(n-1)\rho_a}{nk} \sigma_a^2$ e $\sigma_{c_{5a}} = \frac{1+(nk+n-1)\rho_a}{nk(k+1)} \sigma_a^2$. No presente trabalho, os valores de $\sigma_{c_{4a}}$ e $\sigma_{c_{5a}}$ foram equivalentes a 0,3125 e 0,1875 da variância genética aditiva, respectivamente. Considerando um látice balanceado com $k^2 = 64$ tratamentos, 9 repetições e 6 plantas por parcela, estes valores corresponderiam a 0,0468 e 0,0330 da variância genética aditiva, respectivamente, resultados estes que pouco contribuiriam para um aumento da acurácia pelo ajuste do efeito de blocos dentro de repetições como aleatório. Este resultado foi confirmado por estudos de simulação realizados por Ugarte et al. (1992) e Visscher & Goddard (1993) que concluíram que o modelo do tipo BLUP (2) é mais vantajoso que aquele do tipo BLUP (1) somente quando cada nível do efeito em consideração for pequeno. Esses autores concluíram também que a vantagem do BLUP (2) sobre o BLUP (1) se verifica quando $\sigma_{rd}^2 / \sigma_b^2$ for de alta magnitude.

A diferença nos valores de acurácia associados ao BLUP (1) e BLUP (2) é devida a dois fatores: (i) aumento na acurácia do BLUP (2) em decorrência do uso adicional do efeito de bloco dentro de repetição e (ii) redução na acurácia do BLUP (1) em função da estimativa do efeito de bloco, a partir de um menor número de observações do que aquele usado na estimativa do efeito de repetição pelo BLUP (2).

A estimativa dos efeitos fixos simultaneamente à predição dos valores genéticos conduz a um acréscimo (em relação ao BLP que supõe os efeitos fixos como conhecidos) na PEV, porque um erro adicional é introduzido em função do uso de uma estimativa destes efeitos e não de seu valor verdadeiro. A PEV pode ser expressa por $PEV = G - C'V^{-1}C + C'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}C$ (em que $G = Var(a)$ e $C = COV(a, y)$) (Henderson, 1984) em que a terceira parte desta expressão refere-se ao acréscimo devido à estimativa dos efeitos fixos. Segundo esta expressão, no presente trabalho, $PEV_1 = 0,8826 + 0,0712 = 0,9539$ e $PEV_2 = 0,8922 + 0,0325 = 0,9247$ para o BLUP (1) e BLUP (2), respectivamente. Verifica-se que a contribuição para a PEV, associada à estimativa dos efeitos fixos, é maior para o BLUP (1) do que para o BLUP (2). Este resultado é devido ao menor número de observações usadas na estimativa das médias de bloco pelo BLUP (1). De maneira geral, à medida que o número de observações em cada nível do efeito fixo aumenta, diminui o acréscimo na PEV (Kennedy & Trus, 1993).

Em resumo, as comparações entre os modelos BLUP (1) e BLUP (2) dependem de três fatores: (i) do tamanho do bloco; (ii) do tipo de alocação dos materiais genéticos aos blocos (aleatória ou não); (iii) da relação $\sigma_{rd}^2 / \sigma_b^2$. Com o aumento do número de tratamentos, tanto a unidade bloco quanto a repetição aumentam, reduzindo a contribuição da estimativa dos efeitos fixos para a PEV tanto para o BLUP (1) quanto para o BLUP (2). Assim, o aumento do tamanho do bloco resulta em menor acréscimo na acurácia pela inclusão do efeito de blocos como aleatório no BLUP (2). Estes resultados tendem a conduzir a acuráncias próximas pelo BLUP (1) e BLUP (2).

O tipo de alocação dos materiais genéticos aos blocos torna-se, então, relevante, embora muitas vezes não esteja sob controle do melhorista. Considerando que o delineamento em látice balanceado propicia uma boa conectabilidade (capacidade para estimar diferenças ambientais entre blocos, conforme Weeks & Williams, 1964), a adoção do modelo de blocos fixos permite, com segurança, contornar este possível fator indesejável associado à correlação entre efeitos de blocos e de materiais genéticos. No presente caso, o bloco 1 encontra-se conectado aos blocos 3 e 5 em função do tratamento 1 e, aos blocos 4 e 6, por meio do tratamento 2. Assim, tal bloco não se encontra conectado apenas ao bloco 2, permitindo, desta forma, um uso eficiente do modelo BLUP (1).

Análise comparativa entre os delineamentos de blocos ao acaso e látice, para propósitos de seleção

Na Tabela 7.24 são apresentados os resultados referentes ao índice I_3 e procedimento BLUP(3) associados ao delineamento de blocos ao acaso.

Tabela 7.24. Coeficientes de ponderação (b) das diferentes informações fenotípicas, valores genéticos preditos, variâncias dos erros de predição (PEV) e acuráncias ($r_{\hat{aa}_b}$) associadas ao índice I_3 e procedimento BLUP(3).

b_i	Indivíduos	Valores genéticos preditos	PEV	$r_{\hat{aa}_b}$
$b_1=0,1392$	1	-0,6665	0,9344	0,3766
$b_6=0,1008$	2	-0,6484	0,9344	0,3766
$b_7=0,2516$	3	0,6157	0,9344	0,3766
	4	0,6881	0,9344	0,3766
	5	-0,5702	0,9344	0,3766
	6	-0,5576	0,9344	0,3766
	7	0,5945	0,9344	0,3766
	8	0,5444	0,9344	0,3766
	9	-0,6821	0,9344	0,3766
	10	-0,6195	0,9344	0,3766

Continua...

Tabela 7.24. Continuação.

b_i	Indivíduos	Valores genéticos preditos	PEV	$r_{\hat{aa}_b}$
	11	-0,6366	0,9344	0,3766
	12	-0,6421	0,9344	0,3766
	13	0,6803	0,9344	0,3766
	14	0,8293	0,9344	0,3766
	15	0,5012	0,9344	0,3766
	16	0,5694	0,9344	0,3766
	17	-1,0092	0,9344	0,3766
	18	-0,7587	0,9344	0,3766
	19	-0,0865	0,9344	0,3766
	20	0,8461	0,9344	0,3766
	21	0,4264	0,9344	0,3766
	22	0,6351	0,9344	0,3766
	23	0,3909	0,9344	0,3766
	24	-0,4442	0,9344	0,3766
	25*	-0,9743	0,9518	0,3547
	26*	0,8611	0,9518	0,3547
	27*	-0,5466	0,9518	0,3547
	28*	0,6598	0,9518	0,3547

* referente às 4 genitoras.

Verifica-se que a análise considerando o delineamento como blocos casualizados conduziu a maiores acurárias que o BLUP(1), porém, menores que o BLUP(2). Comparando com a análise com recuperação da informação interblocos do látice (que equivale ao BLUP(2)), a eficiência do látice pode ser dada por $r_{\hat{aa}_2} / r_{\hat{aa}_b}$, correspondendo a $0,3882/0,3766=1,03$ para a seleção de indivíduos e $0,3711/0,3547 = 1,05$ para a seleção de parentais. Estas eficiências são mais baixas do que aquela (1,09) calculada com base na relação entre variâncias residuais e mais altas do que aquela (1,01) baseada no erro efetivo do látice. Isto mostra que, para o objetivo de seleção, as eficiências relativas entre os delineamentos podem mudar, conforme Schutz & Cockerham (1966).

Considerações finais

(i) A análise intrablocos do látice não é adequada, pois conduz a médias ajustadas viciadas, mesmo para delineamentos balanceados; (ii) a análise do látice com recuperação da informação interblocos conduz a resultados similares aos obtidos pelo procedimento BLUP ajustando bloco dentro de repetição como efeito aleatório e ambos podem conduzir a comparações genéticas viciadas; (iii) o procedimento BLUP ajustando blocos como efeito fixo conduz a resultados não viciados; (iv) para pequeno número de tratamentos o procedimento BLUP ajustando blocos como efeito aleatório conduziu à maior acurácia seletiva e menor variância do erro de predição dos valores genéticos; (v) para experimentos

com maior (mais de 60 tratamentos) número de tratamentos (blocos grandes), os dois modelos BLUP tendem a produzir acurárias similares, (vi), genericamente os efeitos de blocos incompletos podem ser tratados como aleatórios, sem prejuízo.

7.9. BLUP Individual e Índice Multiefeitos Envolvendo Diferentes Populações

A experimentação com progênies de diferentes populações (testes de progénie/procedência) é prática comum no melhoramento florestal e em outras espécies perenes. A seleção nesta situação deve basear-se em modelos que incluem o efeito de população ou procedência, o qual pode ser considerado fixo (geralmente quando o número de populações é pequeno) ou aleatório.

A consideração do efeito de procedência como fixo (White & Hodge, 1989; Resende et al., 1996a e b) no procedimento BLUP implica apenas o ajuste desse efeito em β (vetor de efeitos fixos), de forma similar aos ajustes realizados para efeitos de blocos, sendo as equações de modelo misto para os efeitos aleatórios (valores genéticos e parcelas) idênticas às apresentadas para a seleção que envolve apenas uma população. Neste caso, a solução para os valores genéticos dos indivíduos equivale à soma dos valores genéticos preditos dentro de população com os efeitos de procedências estimados em β . É importante relatar que, quando o efeito de procedências é de fato aleatório, mas, tratado como fixo, assume-se que a variância entre procedências tem magnitude muito maior que as variâncias dos demais efeitos aleatórios, ou seja, considera-se que a herdabilidade da média de procedência tende a 1.

A consideração dos efeitos de procedência como aleatórios demanda desenvolvimentos adicionais, tanto para o caso balanceado (índices multiefeitos), quanto desbalanceado (BLUP individual). A seguir são apresentados os desenvolvimentos realizados por Resende (1999b).

Teste de Procedências e Progênies

Modelos e componentes de variação

O modelo linear associado à avaliação de f famílias dentro de cada uma de p procedências, no delineamento de blocos ao acaso com b repetições e n plantas por parcela é dado por:

$$Y_{ijkl} = \mu + p_i + f_{ij} + b_k + pb_{ik} + e_{ijk} + \delta_{ijkl}, \text{ em que:}$$

Y_{ijkl} : observação do indivíduo $ijkl$.

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$.

p_i : efeito da procedência ou população i , aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$.

f_{ij} : efeito da família j , dentro da população i , aleatório, $E(f_{ij})=0$ e $E(f_{ij}^2)=\sigma_f^2$.

b_k : efeito fixo do bloco ou repetição k , $E(b_k)=b_k$ e $E(b_k^2)=V_b$.

pb_{ik} : efeito da interação da procedência i com o bloco k , aleatório, $E(pb_{ik})=0$ e $E(pb_{ik}^2)=\sigma_{pb}^2$.

e_{ijk} : efeito da interação do bloco k com a família j da procedência i , aleatório, $E(e_{ijk})=0$ e $E(e_{ijk}^2)=\sigma_e^2$.

δ_{ijkl} : efeito do indivíduo ℓ dentro da parcela ijk , aleatório, $E(\delta_{ijkl})=0$ e $E(\delta_{ijkl}^2)=\sigma_\delta^2$.

Os componentes de variância apresentados têm os seguintes significados:

σ_p^2 , e σ_f^2 : variâncias entre procedências e entre famílias dentro de procedências, respectivamente.

$\sigma_e^2 = c_1^2 \sigma_F^2$: variância entre parcelas de famílias.

$\sigma_{pb}^2 = c_2^2 \sigma_F^2$: variância entre parcelas de procedências.

σ_δ^2 : variância dentro de parcelas de famílias, em que:

$\sigma_y^2 = \sigma_p^2 + \sigma_f^2 + \sigma_{pb}^2 + \sigma_e^2 + \sigma_\delta^2$: variância fenotípica individual.

$h^2 = \frac{\sigma_a^2}{\sigma_y^2}$: herdabilidade individual no sentido restrito.

$c_1^2 = \frac{\sigma_e^2}{\sigma_y^2}$: fração da correlação intraclasse entre os indivíduos da parcela, de vida ao ambiente comum.

$c_2^2 = \frac{\sigma_{pb}^2}{\sigma_y^2}$: correlação intraclasse entre os indivíduos de diferentes famílias, mas da mesma procedência em um mesmo bloco.

$\sigma_a^2 = \sigma_{aD}^2 + \sigma_{aE}^2$: somatório da variação genética aditiva dentro e entre procedências.

$\sigma_{aD}^2 = \sigma_f^2 / \rho_a$: variância genética aditiva dentro de procedências.

$\sigma_{aE}^2 = \sigma_p^2$ variância genética aditiva entre procedências.

ρ_a : correlação genética intraclasse entre os indivíduos de uma família ($\rho_a = 0,25$ para meios-irmãos e apresenta valores variados para progêneres de polinização aberta de acordo com a taxa de autofecundação, conforme Resende et al., 1995c).

Em termos da variação aditiva total (σ_a^2), a variância entre procedências é dada por: $\sigma_{aE}^2 = \rho_p \sigma_a^2$, em que: $\rho_p = \frac{COV_a(Y_{ijkl}, Y_{ijk'l'})}{\sigma_a^2} = \frac{\sigma_{aE}^2}{\sigma_a^2} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_f^2 / \rho_a} =$ correlação genética intraclasse entre indivíduos da mesma procedência mas de diferentes famílias. Por outro lado, em termos de variação aditiva total, a variância entre progênies dentro de procedências é dada por $\sigma_f^2 = \rho_a (1 - \rho_p) \sigma_a^2$ e a variância aditiva dentro de procedências é dada por $\sigma_{aD}^2 = \sigma_a^2 - \sigma_p^2 = \sigma_a^2 - \rho_p \sigma_a^2 = (1 - \rho_p) \sigma_a^2$.

Índice multiefeitos para a predição de efeitos aditivos

O índice multiefeitos incluindo o efeito de procedências corresponde a:

$$\begin{aligned} I_1 &= b_1 d_{ijkl} + b_2 e_{ijk} + b_3 f_{ij} + b_4 p_i + b_5 pb_{ik} \\ &= b_1 (Y_{ijkl} - \bar{Y}_{ijk.}) + b_2 (\bar{Y}_{ijk.} - \bar{Y}_{ij..} - \bar{Y}_{i.k.} + \bar{Y}_{i...}) + b_3 (\bar{Y}_{ij..} - \bar{Y}_{i...}) + b_4 (\bar{Y}_{i...} - \bar{Y}_{....}) + \\ &\quad + b_5 (\bar{Y}_{i.k.} - \bar{Y}_{i...} - \bar{Y}_{..k.} + \bar{Y}_{....}), \text{ em que:} \end{aligned}$$

$$b_1 = \frac{(1 - \rho_a) \sigma_{aD}^2}{\sigma_\delta^2} = \frac{(1 - \rho_a) (1 - \rho_p) \sigma_a^2}{\sigma_\delta^2} = \frac{(1 - \rho_a) (1 - \rho_p) h^2}{(1 - t_1 - c_2^2)}$$

$$b_2 = \frac{[(1 - \rho_f)/n] \sigma_{aD}^2}{\frac{\sigma_\delta^2 + \sigma_e^2}{n}} = \frac{\{[(1 - \rho_a) (1 - \rho_p)]/n\} \sigma_a^2}{\frac{\sigma_\delta^2 + \sigma_e^2}{n}} = \frac{(1 - \rho_a) (1 - \rho_p) h^2}{1 + (n-1) t_1 - nt_2 - c_2^2}$$

$$\begin{aligned} b_3 &= \frac{\frac{1 + (nb-1) \rho_a}{nb} \sigma_{aD}^2}{\frac{\sigma_f^2 + \frac{\sigma_e^2}{b} + \frac{\sigma_\delta^2}{nb}}{\sigma_f^2 + \frac{\sigma_e^2}{b} + \frac{\sigma_\delta^2}{nb}}} = \frac{\left[(1 - \rho_p) \rho_a + \frac{(1 - \rho_a) (1 - \rho_p)}{nb} \right] \sigma_a^2}{\sigma_f^2 + \frac{\sigma_e^2}{b} + \frac{\sigma_\delta^2}{nb}} = \\ &= \frac{\left[nb (1 - \rho_p) \rho_a + (1 - \rho_a) (1 - \rho_p) \right] h^2}{1 + (n-1) t_1 + (b-1) nt_2 - nb t_3 - c_2^2} \end{aligned}$$

$$\begin{aligned}
b_4 &= \frac{\sigma_{aE}^2 + \frac{1+(nb-1) \rho_a}{nfb} \sigma_{aD}^2}{\sigma_p^2 + \frac{\sigma_f^2}{f} + \frac{\sigma_e^2}{fb} + \frac{\sigma_{pb}^2}{b} + \frac{\sigma_\delta^2}{nfb}} = \frac{\rho_p \sigma_a^2 + \frac{(1-\rho_f) (1-\rho_p)}{nfb} \sigma_a^2 + \frac{(1-\rho_p) \rho_a}{f} \sigma_a^2}{\sigma_p^2 + \frac{\sigma_f^2}{f} + \frac{\sigma_e^2}{fb} + \frac{\sigma_{pb}^2}{b} + \frac{\sigma_\delta^2}{nfb}} = \\
&= \frac{[nfb \rho_p + (1-\rho_a) (1-\rho_p) + nb (1-\rho_p) \rho_a] h^2}{1 + (n-1) t_1 + (b-1) nt_2 + (f-1) nbt_3 + (nf-1) c_2^2} \\
b_5 &= \frac{[(1-\rho_a)/nf] \sigma_{aD}^2}{\frac{\sigma_\delta^2}{nf} + \frac{\sigma_e^2}{f} + \sigma_{pb}^2} = \frac{\{[(1-\rho_a) (1-\rho_p)]/nf\} \sigma_a^2}{\frac{\sigma_\delta^2}{nf} + \frac{\sigma_e^2}{f} + \sigma_{pb}^2} = \frac{(1-\rho_a) (1-\rho_p) h^2}{1 + (n-1) t_1 - nt_2 + (nf-1) c_2^2} \\
t_2 &= \frac{\rho_a \sigma_{aD}^2 + \sigma_{aE}^2}{\sigma_y^2} = \frac{[\rho_a (1-\rho_p) + \rho_p] \sigma_a^2}{\sigma_y^2} = [\rho_a (1-\rho_p) + \rho_p] h^2: \quad \text{correlação intraclass entre indivíduos de uma família em diferentes blocos.} \\
t_1 = t_2 + c_1^2 &: \quad \text{correlação intraclass entre indivíduos de uma mesma família no mesmo bloco.} \\
t_3 &= \frac{\sigma_{aE}^2}{\sigma_y^2} = \frac{\rho_p \sigma_a^2}{\sigma_y^2} = \rho_p h^2 = h_E^2: \quad \text{correlação intraclass entre indivíduos de diferentes famílias mas da mesma procedência.}
\end{aligned}$$

Considerando as expressões apresentadas em termos da herdabilidade e dos coeficientes de correlação intraclass, verifica-se que, para a avaliação genética, são necessárias estimativas apenas dos parâmetros h^2 , c_1^2 e c_2^2 , das quais h^2 é dada pela soma das herdabilidades ao nível de indivíduo dentro de procedência ($h_D^2 = \sigma_{aD}^2 / \sigma_y^2$) e entre procedências ($h_E^2 = \sigma_{aE}^2 / \sigma_y^2 = t_3$). O parâmetro ρ_p é dado por $\rho_p = h_E^2 / h^2$.

Os componentes dos efeitos do índice são:

\bar{Y}_{ijkl} , $\bar{Y}_{ijk.}$, $\bar{Y}_{ij..}$ e $\bar{Y}_{...}$: valor individual, média da parcela, média da família e média da procedência, respectivamente.

$\bar{Y}_{i.k.}$, $\bar{Y}_{..k.}$ e $\bar{Y}_{...}$: média da procedência no bloco, média do bloco e média geral, respectivamente.

Índice multiefeitos com medidas repetidas para a predição dos efeitos aditivos

Considerando a realização de várias medições (safras) em cada indivíduo, o índice multiefeitos, incluindo o efeito de população é:

$I_2 = b_6 \delta_{ijkl}^* + b_7 e_{ijk} + b_8 f_{ij} + b_9 p_i + b_{10} pb_{ik}$, em que:

δ_{ijkl}^* = efeito individual (dentro de parcelas) médio de várias safras.

$$b_6 = \frac{(1-\rho_a)(1-\rho_p) h^2}{[1+(m-1)\rho - mt_1 - c_2^2]/m};$$

$$b_7 = \frac{(1-\rho_a)(1-\rho_p) h^2}{[1+(m-1)\rho + (n-1)mt_1 - nmt_2 - c_2^2]/m};$$

$$b_8 = \frac{[nb(1-\rho_p)\rho_a + (1-\rho_a)(1-\rho_p)]h^2}{[1+(m-1)\rho + (n-1)mt_1 + (b-1)nmt_2 - nbmt_3 - c_2^2]/m};$$

$$b_9 = \frac{[nfb\rho_p + (1-\rho_a)(1-\rho_p) + nb(1-\rho_p)\rho_a]h^2}{[1+(m-1)\rho + (n-1)mt_1 + (b-1)nmt_2 + (f-1)nbmt_3 + (nfm-1)c_2^2]/m};$$

$$b_{10} = \frac{(1-\rho_a)(1-\rho_p)h^2}{[1+(m-1)\rho + (n-1)mt_1 - nmt_2 + (nfm-1)c_2^2]/m}, \text{ em que:}$$

$$\rho = \frac{\sigma_{ad}^2 + \sigma_{ae}^2 + \sigma_d^2 + \sigma_{ep}^2 + \sigma_e^2}{\sigma_y^2} = \frac{\sigma_f^2/\rho_a + \sigma_p^2 + (\sigma_\delta^2 - \sigma_f^2/\rho_a + \sigma_f^2 - \sigma_{et}^2) + \sigma_e^2}{\sigma_y^2} = \frac{\sigma_f^2 + \sigma_p^2 + \sigma_\delta^2 + \sigma_e^2}{\sigma_y^2};$$

repetibilidade ao nível de indivíduo.

σ_d^2 , σ_{ep}^2 e σ_{et}^2 : variância genética de dominância, ambiental permanente e ambiental temporária, respectivamente.

$\sigma_y^2 = \sigma_{ad}^2 + \sigma_{ae}^2 + \sigma_d^2 + \sigma_{ep}^2 + \sigma_e^2 + \sigma_{pb}^2 + \sigma_{et}^2$: variância fenotípica ao nível de indivíduo.

7.9.1. BLUP individual para a predição de efeitos aditivos

O modelo linear misto para a avaliação genética ao nível de indivíduo é dado por:

$$y = X\beta + Za + Wc_1 + Tc_2 + e, \text{ em que:}$$

y , β , a , c_1 , c_2 e e : vetores de dados ao nível de indivíduo, de efeitos fixos (médias de blocos), de efeitos genéticos aditivos (aleatório), de efeitos de parcelas de progénie (aleatório), de efeitos de parcela de procedência (aleatório) e de erros aleatórios, respectivamente.

X , Z , W e T : matrizes de incidência para β , a , c_1 e c , respectivamente.

As equações de modelo misto para predição dos efeitos fixos e aleatórios equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W & Z'T & Z'Q \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'Q \\ T'X & T'Z & T'W & T'T + I\lambda_3 & T'Q \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \end{bmatrix} \quad (1), \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_{ad}^2}{\sigma_a^2} = \frac{1-h^2-c_1^2-c_2^2}{h^2}; \quad \lambda_2 = \frac{\sigma_{ad}^2}{\sigma_e^2} = \frac{1-h^2-c_1^2-c_2^2}{c_1^2}; \quad \lambda_3 = \frac{\sigma_{ad}^2}{\sigma_{pb}^2} = \frac{1-h^2-c_1^2-c_2^2}{c_2^2}.$$

A : matriz de parentesco genético aditivo entre os indivíduos sob avaliação;
 $\sigma_{ad}^2 = \sigma_\delta^2 - (1-\rho_a)(1-\rho_p)\sigma_a^2$.

Para a avaliação genética que envolve medidas repetidas o modelo linear misto equivale a:

$y = X\beta + Za + Wc_1 + Tc_2 + Qp + e$, em que p refere-se aos efeitos permanentes (de dominância + de ambiente permanente) e Q é a matriz de incidência para p .

As equações de modelo misto são:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{c}_1 \\ \hat{c}_2 \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q \\ Z'X & Z'Z + A^{-1}\lambda_4 & Z'W & Z'T & Z'Q \\ W'X & W'Z & W'W + I\lambda_5 & W'T & W'Q \\ T'X & T'Z & T'W & T'T + I\lambda_6 & T'Q \\ Q'X & Q'Z & Q'W & Q'T & Q'Q + I\lambda_7 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ Q'y \end{bmatrix} \quad (2), \text{ em que:}$$

$$\lambda_4 = \frac{\sigma_{et}^2}{\sigma_a^2} = \frac{1-\rho-c_2^2}{h^2}; \quad \lambda_5 = \frac{\sigma_{et}^2}{\sigma_e^2} = \frac{1-\rho-c_2^2}{c_1^2};$$

$$\lambda_6 = \frac{\sigma_{et}^2}{\sigma_{pb}^2} = \frac{1-\rho-c_2^2}{c_2^2}; \quad \lambda_7 = \frac{\sigma_{et}^2}{\sigma_{ep}^2 + \sigma_D^2} = \frac{1-\rho-c_2^2}{\rho-h^2-c_1^2}$$

7.9.2. Teste de procedências ou populações

Em muitas situações práticas de campo, são instalados testes de procedências em que cada população é representada por uma mistura de sementes de várias matrizes, não havendo assim a identificação de progêneres. Estes tipos de testes são importantes no melhoramento de árvores não industriais e de uso múltiplo (Haugerud & Collinson, 1990; Simons, 1992). Os experimentos são posteriormente transformados em pomares de sementes por mudas (Nanson, 1972). A técnica do índice multiefeitos pode também ser utilizada nesta situação conforme especificado a seguir.

Modelo estatístico e componentes de variância

O modelo linear associado à avaliação de p procedências no delineamento de blocos ao acaso com b repetições e n plantas por parcela equivale a:

$$Y_{ijk\ell} = \mu + p_i + b_j + e_{ij} + \delta_{ijk}, \text{ em que:}$$

Y_{ijk} : observação do indivíduo ijk ;

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$;

p_i : efeito da procedência i , aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$;

b_j : efeito do bloco j , fixo, $E(b_j) = b_J$ e $E(b_j^2) = V_b$;

e_{ij} : efeito da parcela ij , aleatório, $E(e_{ij}) = 0$ e $E(e_{ij}^2) = \sigma_e^2$;

δ_{ijk} : efeito do indivíduo k dentro da parcela ij , aleatório, $E(\delta_{ijk}) = 0$ e $E(\delta_{ijk}^2) = \sigma_\delta^2$.

Os seguintes parâmetros genéticos e fenotípicos estão associados ao modelo:

$h_E^2 = \frac{\sigma_p^2}{\sigma_y^2} = t_3$: "herdabilidade" entre procedências ao nível de indivíduo ou correlação intraclasse entre indivíduos de uma mesma procedência.

$c^2 = \frac{\sigma_e^2}{\sigma_y^2}$: correlação intraclasse devida ao ambiente comum da parcela.

$\sigma_y^2 = \sigma_p^2 + \sigma_e^2 + \sigma_\delta^2$: variância fenotípica.

Considerando toda a variação genética aditiva entre e dentro de populações, define-se:

$$\sigma_a^2 = \sigma_{aE}^2 + \sigma_{aD}^2; \quad h^2 = h_E^2 + h_D^2.$$

Índice multiefeitos para efeitos aditivos

Neste caso o índice multiefeitos é:

$$I_3 = b_{11} \delta_{ijk} + b_{12} e_{ij} + b_{13} p_i$$

$= b_{11} (Y_{ijk} - \bar{Y}_{ij.}) + b_{12} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}) + b_{13} (\bar{Y}_{i..} - \bar{Y}_{...})$, em que:

$$b_{11} = \frac{\sigma_{aD}^2}{\sigma_\delta^2} = \frac{(1-\rho_p)\sigma_a^2}{\sigma_\delta^2} = \frac{(1-\rho_p) h^2}{1-h_E^2-c^2} = \frac{(1-\rho_p) h^2}{1-\rho_p h^2-c^2}$$

$$\begin{aligned}
 b_{12} &= \frac{\sigma_{ad}^2 / n}{\frac{\sigma_\delta^2}{n} + \sigma_e^2} = \frac{(1 - \rho_p) h^2}{1 + (n-1) c^2 - \rho_p h^2} \\
 b_{13} &= \frac{\sigma_p^2 + \sigma_{ad}^2 / nb}{\sigma_p^2 + \frac{\sigma_e^2}{b} + \frac{\sigma_\delta^2}{nb}} = \frac{nb h_E^2 + h_D^2}{1 + (n-1) (h_E^2 + c^2) + (b-1) nh_E^2} = \\
 &= \frac{[1 + (nb-1)\rho_p] h^2}{1 + (n-1) (\rho_p h^2 + c^2) + (b-1) n\rho_p h^2}
 \end{aligned}$$

Assim, para uma avaliação genética completa, são necessários os parâmetros h_E^2 , h_D^2 , c^2 e ρ_p , dos quais h_E^2 e c^2 podem ser calculados dos próprios dados experimentais. É preciso que se conheça, previamente, h_D^2 (obtido de outro experimento com estrutura de progênie ou de literatura). O parâmetro ρ_p neste caso é dado por $\rho_p = h_E^2 / h^2$.

Índice multiefeitos com medidas repetidas para a predição de efeitos aditivos

Considerando a realização de várias medições (safras) em cada indivíduo, o índice multiefeitos é obtido por:

$$I_4 = b_{14} \delta_{ijk}^* + b_{15} e_{ij} + b_{16} p_i, \text{ em que}$$

δ_{ijk}^* : valor individual médio de várias safras.

$$b_{14} = \frac{(1 - \rho_p) h^2}{[1 + (m-1) \rho - m\rho_p h^2 - mc^2]/m}$$

$$b_{15} = \frac{(1 - \rho_p) h^2}{[1 + (m-1) \rho + (n-1) mc^2 - m\rho_p h^2]/m}$$

$$b_{16} = \frac{[1 + (nb-1) \rho_p] h^2}{[1 + (m-1) \rho + (n-1) m (\rho_p h^2 + c^2) + (b-1) nm \rho_p h^2]/m}, \text{ em que:}$$

$$\rho = \frac{\sigma_{ad}^2 + \sigma_{ae}^2 + \sigma_d^2 + \sigma_{ep}^2 + \sigma_e^2}{\sigma_y^2} = \frac{\sigma_p^2 + (\sigma_\delta^2 - \sigma_{ei}^2) + \sigma_e^2}{\sigma_y^2} = h_E^2 + \rho_D + c^2 : \text{repetibilidade}$$

ao nível de indivíduo;

$$\rho_D = \frac{\sigma_{ad}^2 + \sigma_d^2 + \sigma_{ep}^2}{\sigma_y^2} : \text{repetibilidade ao nível de indivíduo dentro de procedência.}$$

BLUP individual para a predição de efeitos aditivos

Neste caso, o modelo linear misto corresponde a:

$$y = X\beta + Za + Wc + e, \text{ em que:}$$

y, β, a, c e e : vetores de dados ao nível de indivíduo, de efeitos fixos (médias de blocos), de efeitos genéticos aditivos (aleatório), de efeitos de parcela (aleatório) e de erros aleatórios, respectivamente.

As equações de modelo misto equivalem a:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_8 & Z'W \\ W'X & W'Z & W'W + I\lambda_9 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix} \quad (3), \text{ em que:}$$

$$\lambda_8 = \frac{\sigma_r^2}{\sigma_a^2} = \frac{1-h^2-c^2}{h^2}; \quad \lambda_9 = \frac{\sigma_r^2}{\sigma_e^2} = \frac{1-h^2-c^2}{c^2}; \quad \sigma_r^2: \text{ variância residual dentro de parcela.}$$

Para a avaliação genética que envolve medidas repetidas, o modelo linear misto equivale a: $y = X\beta + Za + Wc + Tp + e$, em que p refere-se aos efeitos permanentes (de dominância + de ambiente permanente) e T é a matriz de incidência para p .

As equações de modelo misto são:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'T \\ Z'X & Z'Z + A^{-1}\lambda_{10} & Z'W & Z'T \\ W'X & W'Z & W'W + I\lambda_{11} & W'T \\ T'X & T'Z & T'W & T'T + I\lambda_{12} \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \end{bmatrix} \quad (4), \text{ em que:}$$

$$\lambda_{10} = \frac{\sigma_{et}^2}{\sigma_a^2} = \frac{1-\rho}{h^2}; \quad \lambda_{11} = \frac{\sigma_{et}^2}{\sigma_e^2} = \frac{1-\rho}{c^2}; \quad \lambda_{12} = \frac{\sigma_{et}^2}{\sigma_{ep}^2 + \sigma_D^2} = \frac{1-\rho}{\rho - h^2 - c^2}$$

7.9.3. Exemplo prático

Para exercício das metodologias, foram considerados os seguintes parâmetros para teste de procedências e progênies com uma medição por indivíduo:

$$h^2 = 0,25; \quad h_E^2 = 0,025; \quad \rho_p = 0,10; \quad \rho_a = 0,25; \quad c^2 = 0,10; \quad c_1^2 = 0,05; \quad c_2^2 = 0,05;$$

$$\lambda_1 = 2,4; \quad \lambda_2 = 6,0; \quad \lambda_3 = 12,0$$

Para o teste de procedência com uma medição por indivíduo, foram considerados os parâmetros:

$$h^2 = 0,25; \quad h_E^2 = 0,025; \quad \rho_p = 0,10; \quad c^2 = 0,05; \quad \lambda_8 = 2,8; \quad \lambda_9 = 14,0; \quad \rho_p = 0,10.$$

Para aplicação das metodologias, considerou-se o seguinte subconjunto de dados de *Eucalyptus grandis*.

Indivíduos	Bloco	Procedência	Progênie	Parcela	Diâmetro (cm)
1	1	1	1	1	9,87
2	1	1	1	1	10,00
3	1	1	2	2	14,48
4	1	1	2	2	15,00
5	1	2	3	3	8,91
6	1	2	3	3	9,00
7	1	2	4	4	15,00
8	1	2	4	4	14,64
9	2	1	1	5	9,55
10	2	1	1	5	10,00
11	2	1	2	6	8,00
12	2	1	2	6	7,96
13	2	2	3	7	15,00
14	2	2	3	7	16,07
15	2	2	4	8	14,01
16	2	2	4	8	14,50

Os vetores e matrizes apresentam as seguintes dimensões neste exemplo:

$y = (16 \times 1)$; $\beta = (2 \times 1)$; $a = (22 \times 1)$ incluindo as quatro matrizes parentais e as 2 procedências; $c_1 = (8 \times 1)$; $c_2 = (4 \times 1)$; $X = (16 \times 2)$; $Z = (22 \times 22)$; $W = (16 \times 8)$; $T = (16 \times 4)$.

A matriz A (22×22) apresenta valores 1 na diagonal, 0,325 (dado por $\rho_a(1 - \rho_p) + \rho_p$) nas células correspondentes aos parentescos de indivíduos de uma mesma progênie, 0,10 nas células relativas aos parentescos entre indivíduos de uma mesma procedência, 0,55 nas células correspondentes aos parâmetros mãe – filho (dado por $0,5(1 - \rho_p) + \rho_p$), 0,10 nas células correspondentes aos parentescos procedência e genitor e 0 nas demais células.

Para o teste de procedências considerou-se o seguinte conjunto de dados:

Indivíduos	Bloco	Procedência	Parcela	Diâmetro (cm)
1	1	1	1	9,87
2	1	1	1	14,48
3	1	2	2	8,91
4	1	2	2	15,00
5	2	1	3	9,55
6	2	1	3	8,00
7	2	2	4	15,00
8	2	2	4	14,01

Neste caso, os vetores e matrizes têm as seguintes dimensões:

$y = (8 \times 1)$; $\beta = (2 \times 1)$; $a = (10 \times 1)$ incluindo as 2 procedências; $c = (4 \times 1)$; $X = (8 \times 2)$; $Z = (8 \times 8)$; $W = (8 \times 4)$.

A matriz A (10×10) apresenta valores 1 na diagonal, 0,10 nas células relativas aos parentescos entre indivíduos de uma mesma procedência, 0,10 nos parentescos entre indivíduos de uma procedência e a respectiva procedência e 0 nas demais células.

Os resultados referentes ao índice I_1 e método BLUP (1) são apresentados na Tabela 7.25.

Tabela 7.25. Coeficientes de ponderação (b_i) dos diferentes efeitos do modelo estatístico e efeitos aditivos preditos pelo índice I_1 e procedimento BLUP(1), bem como demais efeitos aleatórios preditos pelo BLUP(1).

b_i	Indivíduos	Efeitos aditivos	\hat{c}_1	\hat{c}_2
$b_1 = 0,21951$	1	-0,8350	-0,3647	0,1020
$b_2 = 0,17419$	2	-0,8064	0,5687	-0,1020
$b_3 = 0,32984$	3	0,1935	-0,6628	-0,4514
$b_4 = 0,37255$	4	0,3076	0,4588	0,4514
$b_5 = 0,14439$	5	-0,4140	-0,2368	
	6	-0,3943	-0,6659	
	7	1,0138	0,6281	
	8	0,9348	0,2746	
	9	-0,7622		
	10	-0,6634		
	11	-0,7868		
	12	-0,7956		
	13	0,5676		
	14	0,8025		
	15	0,7651		
	16	0,8727		

O procedimento BLUP(1) propicia também a predição dos efeitos genéticos aditivos para os genitores (matrizes) e para as procedências, os quais são apresentadas na Tabela 7.26.

Tabela 7.26. Coeficientes de ponderação (b_i) e efeitos aditivos preditos pelo BLUP (1) para os genitores (\hat{a}_m) e para procedências (\hat{a}_p).

b_1	\hat{a}_m	\hat{a}_p
$b_{17} = 0,37696$	$\hat{a}_{m_1} = -0,8513$	$\hat{a}_{p_1} = -0,1747$
$b_{18} = 0,28235$	$\hat{a}_{m_2} = -0,2840$	$\hat{a}_{p_2} = 0,1747$
$b_{19} = 0,12549$	$\hat{a}_{m_3} = 0,1356$	
	$\hat{a}_{m_4} = 0,9898$	

No teste de procedência e progénie, os efeitos aditivos foram preditos pelo BLUP(1) que fornece os mesmos resultados que os índices

$I_5 = b_{17}(\bar{Y}_{ij..} - \bar{Y}_{i...}) + b_{18}(\bar{Y}_{i...} - \bar{Y}_{....}) + b_{19}(\bar{Y}_{i...} - \bar{Y}_{....})$ e $I_6 = b_{19}(\bar{Y}_{i...} - \bar{Y}_{....})$ para as matrizes e procedências, respectivamente. Os ponderadores dos índices são dados por:

$$b_{17} = \frac{nb}{1+(n-1)} \frac{0,5 (1-\rho_p) h^2}{t_1 + (b-1) nt_2 - nbt_3};$$

$$b_{18} = \frac{nb}{1+(n-1)} \frac{0,5 (1-\rho_p) h^2}{t_1 + (b-1) nt_2 + (f-1) nbt_3 + (nf-1) c_2^2}$$

$$b_{19} = \frac{nb}{1+(n-1)} \frac{\rho_p h^2}{t_1 + (b-1) nt_2 + (f-1) nbt_3 + (nf-1) c_2^2}$$

É importante relatar que os valores genéticos preditos pelo BLUP(1) para indivíduos e genitores já incluem o valor genético predito para procedência. Esta abordagem é similar àquela em que se considera o efeito de procedência como efeito de grupo genético conforme Westell et al. (1988) e Resende & Rosa-Perez (1999a). Quando o efeito de procedência é considerado fixo, faz-se necessário, posteriormente, somar os valores genéticos preditos de indivíduos dentro de procedência aos efeitos das respectivas procedências.

Os resultados referentes ao índice I_3 e método BLUP(3) são apresentados na Tabela 7.27.

Tabela 7.27. Coeficientes de ponderação (b_i) dos diferentes efeitos do modelo estatístico e efeitos aditivos preditos pelo índice I_3 e procedimento BLUP(3), bem como demais efeitos aleatórios preditos pelo BLUP(3).

b_1	Indivíduos	Efeitos aditivos	\hat{c}
$b_{11} = 0,24324$	1	-0,6321	0,0227
$b_{12} = 0,21951$	2	0,4893	-0,0227
$b_{13} = 0,28889$	3	-0,6693	-0,2676
	4	0,8121	0,2676
	5	-0,5360	
	6	-0,9130	
	7	0,8449	
	8	0,6041	

O procedimento BLUP(3) fornece também as previsões dos efeitos genéticos aditivos de procedência. Neste exemplo, as previsões equivaleram a $\hat{a}_{p_1} = -0,1224$ e $\hat{a}_{p_2} = 0,1224$. Neste caso, o BLUP (3) forneceu resultados idênticos aos fornecidos pelo índice $I_7 = b_{20}(\bar{Y}_{..} - \bar{Y}_{...})$, em que

$$b_{20} = \frac{nb \rho_p h^2}{1 + (n-1)(\rho_p h^2 + c^2) + (b-1)n\rho_p h^2} = 0,08889$$

É importante ressaltar que os desenvolvimentos realizados neste tópico consideram que toda a variação genética entre procedências é de natureza aditiva, ignorando-se a dominância e variância devida ao desequilíbrio de ligação.

Considerações Finais

(i) Os índices e equações de modelo misto apresentadas fornecem as previsões dos efeitos genéticos aditivos de indivíduos, genitores e procedências (populações); (ii) para pleno uso dos procedimentos apresentados são necessárias estimativas dos parâmetros herdabilidade dentro de procedência, “herdabilidade” (correlação intraclasse) entre procedências e correlações devidas ao ambiente comum da parcela de progénie e da parcela de procedência.

7.10. BLUP Individual e Índice Multiefeitos na Seleção Simultânea de Caracteres

Os modelos multivariados destinam-se à avaliação de indivíduos, simultaneamente, para dois ou mais caracteres e apresentam grande relevância no contexto da seleção envolvendo agregados genotípicos como objetivo da seleção. Tais modelos exploram as correlações genéticas e fenotípicas entre caracteres e foram aplicados inicialmente por Henderson & Quaas (1976).

Os modelos multivariados podem também ser aplicados vantajosamente quando o objetivo da seleção é o melhoramento de um único caráter, nas seguintes situações: (i) quando o caráter é avaliado mais de uma vez no indivíduo, mas a correlação genética entre as medições não equivale a 1 (neste caso o modelo multivariado é superior ao modelo de repetibilidade); (ii) quando o caráter é avaliado em parentes em diferentes ambientes e a correlação genética entre a performance nos dois ambientes não é 1; (iii) no contexto da seleção empregando caracteres auxiliares no melhoramento, onde se explora a correlação genética e fenotípica, bem como a diferença entre as herdabilidades dos caracteres.

Uma das principais vantagens da utilização dos modelos multivariados consiste no aumento da acurácia das avaliações. Na seleção usando caracteres auxiliares, a acurácia é dada por h_m , a raiz quadrada da herdabilidade multivariada. A herdabilidade multivariada é dada por:

$$h_m^2 = h_y^2 \left[1 + \frac{(h_x r_{axy}/h_y - r_{xy})^2}{1 - r_{xy}^2} \right],$$
 onde h_x^2 e h_y^2 referem-se às herdabilidades do caráter auxiliar (x) e do caráter objetivo (y), respectivamente, e r_{axy} e r_{xy} são relativas às correlações genéticas e fenotípicas entre os dois caracteres, respectivamente. Neste caso, a eficiência ou ganho em acurácia com a seleção multivariada é dada pela razão entre as acurárias multivariada e univariada, ou seja:

$$Ef = \frac{h_m}{h_y} = \left[1 + \frac{(h_x r_{axy}/h_y - r_{xy})^2}{1 - r_{xy}^2} \right]^{1/2}$$

Por esta expressão verifica-se:

- (i) Se $h_x/h_y r_{axy} = r_{xy}$, a eficiência será 1 e a análise multivariada será equivalente à univariada.
- (ii) Sendo as herdabilidades iguais ($h_x = h_y$), a eficiência depende apenas da diferença entre as correlações genética e fenotípica. Sendo a correlação fenotípica igual a:

$$r_{xy} = r_{xy} h_x h_y + r_{Exy} [(1 - h_x^2)(1 - h_y^2)]^{1/2}$$

onde r_{Exy} é a correlação ambiental entre os dois caracteres, tem-se que: $(r_{axy} - r_{xy}) = r_{axy} - \{ r_{axy} h_x h_y + r_{Exy} [(1 - h_x^2)(1 - h_y^2)]^{1/2} \}$. Por esta última expressão verifica-se que a maximização (que significa maximizar a eficiência da seleção multivariada) da diferença entre r_{axy} e r_{xy} , só pode se dar por meio da maximização da diferença entre r_{axy} e r_{Exy} . Assim, o ganho em acurácia depende da diferença absoluta entre a correlação genética e a correlação ambiental entre os caracteres.

- (iii) Sendo as herdabilidades iguais ($h_x = h_y$) e as correlações genética e ambiental também iguais ($r_{axy} = r_{Exy}$), a análise multivariada será também equivalente à análise univariada.
- (iv) Para os casos em que a correlação genética for superior à correlação fenotípica, o ganho em acurácia será tanto maior quanto maior a herdabilidade do caráter auxiliar em relação à herdabilidade do caráter objetivo do melhoramento.

Uma desvantagem da análise multivariada ou multicaracterística refere-se ao alto custo computacional. O custo da análise multivariada de n caracteres é maior que o custo de n análises univariadas. Adicionalmente, a análise multivariada requer estimativas fidedignas de correlações genéticas e fenotípicas entre os caracteres (Mrode, 1996).

Quanto ao uso do modelo multivariado em lugar do modelo de repetibilidade, pode-se dizer que o modelo de repetibilidade é muito eficiente quando comparado ao multivariado, obtendo-se perdas de apenas 0% a 5% em eficiência quando a correlação genética entre medições sucessivas é alta (acima de 0,80). Nestes casos, a correlação entre ordenamentos pelo modelo de repetibilidade e pelo modelo multivariado é muito próxima de 1 (Mrode, 1996). O grande benefício do modelo de repetibilidade comparado com o multivariado refere-se à menor demanda computacional e à necessidade de estimativas de poucos parâmetros genéticos.

No contexto dos modelos lineares mistos, o seguinte modelo multivariado para a predição dos efeitos aditivos e de dominância pode ser enunciado:

$y = X\beta + Za + Zd + Wc + e$, em que:

y, β, a, d, c e e : vetores de dados, de efeitos fixos, de efeitos genéticos aditivos (aleatório), de efeitos de dominância (aleatório), de efeitos de parcela (aleatório) e de erros aleatórios, respectivamente.

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

Considerando os dois caracteres (altura e diâmetro), os modelos para cada caráter são:

$$y_1 = X_1\beta_1 + Z_1a_1 + Z_1d_1 + W_1c_1 + e_1 \text{ para altura e}$$

$$y_2 = X_2\beta_2 + Z_2a_2 + Z_2d_2 + W_2c_2 + e_2 \text{ para diâmetro}$$

Estes modelos expressos em notação matricial equivalem a:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Para o modelo bivariado, é a seguinte a estrutura de variâncias:

$$Var = \begin{bmatrix} a_1 \\ a_2 \\ d_1 \\ d_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \sigma_{a_i}^2 A & \sigma_{a_{12}} A & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_{a_{12}} A & \sigma_{d_i}^2 A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{d_1}^2 D & \sigma_{d_{12}} D & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{d_{12}} D & \sigma_{d_2}^2 D & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{e_i}^2 I & \sigma_{e_{12}} I & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{e_{12}} I & \sigma_{e_2}^2 I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ad_i}^2 I & \sigma_{ad_{12}} I \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ad_{12}} I & \sigma_{ad_2}^2 I \end{bmatrix}, \text{ em que:}$$

$\sigma_{a_i}^2$: variância genética aditiva para o caráter i .

$\sigma_{d_i}^2$: variância genética de dominância para o caráter i .

$\sigma_{e_i}^2$: variância ambiental entre parcelas para o caráter i .

σ_{ad}^2 : variância ambiental dentro de parcela para o caráter i .

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

$\sigma_{d_{12}}$: covariância genética de dominância entre os caracteres 1 e 2.

$\sigma_{e_{12}}$: covariância ambiental ao nível de parcela entre os caracteres 1 e 2.

$\sigma_{ad_{12}}$: covariância ambiental dentro de parcela entre os caracteres 1 e 2.

A : matriz de parentesco genético aditivo.

D : matriz de parentesco genético de dominância.

As equações de modelo misto para a predição de efeitos aditivos e de dominância pelo procedimento BLUP sob modelo individual correspondem a:

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

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}; \quad \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix}; \quad \hat{d} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix}$$

$$R^{-1} = R_o^{-1} \otimes I; \quad G^{-1} = G_o^{-1} \otimes A^{-1}; \quad G_d^{-1} = G_{do}^{-1} \otimes D^{-1}; \quad C^{-1} = C_o^{-1} \otimes I$$

$$R_o = \begin{bmatrix} \sigma_{ad_1}^2 & \sigma_{ad_{12}} \\ \sigma_{ad_{12}} & \sigma_{ad_2}^2 \end{bmatrix}; \quad G_o = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad G_{do} = \begin{bmatrix} \sigma_{d_1}^2 & \sigma_{d_{12}} \\ \sigma_{d_{12}} & \sigma_{d_2}^2 \end{bmatrix}; \quad C_o = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma_{e_2}^2 \end{bmatrix}$$

\otimes : denota a operação produto de Kronecker.

Se $\sigma_{ad_{12}} = \sigma_{a_{12}} = \sigma_{d_{12}} = \sigma_{e_{12}} = 0$, as equações reduzem-se a análises univariadas dos dois caracteres, considerados como não correlacionados. Por outro lado, se $\sigma_{ad_{12}} = \sigma_{e_{12}} = 0$, as equações tornam-se equivalentes àquelas descritas no tópico 7.5, para a seleção de um mesmo caráter em diferentes ambientes.

Para o caso balanceado, as equações de modelo misto apresentadas, no seu componente aditivo, estão associadas ao seguinte índice multiefeitos (Resende et al., 1994b):

$$I_1 = b_1(Y_{ijk} - \bar{Y}_{ij.}) + b_2(X_{ijk} - \bar{X}_{ij.}) + b_3(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) + b_4(\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...}) +$$

+ $b_5(\bar{Y}_{i..} - \bar{Y}_{...}) + b_6(\bar{X}_{i..} - \bar{X}_{...})$, para o caráter y , em que:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{n-1}{n} & \sigma_{\delta_i}^2 & \frac{n-1}{n} & \sigma_{\delta_{12}} \\ \frac{n-1}{n} & \sigma_{\delta_{21}} & \frac{n-1}{n} & \sigma_{\delta_2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \frac{n-1}{n} (1-\rho_a) \sigma_{a_i}^2 \\ \frac{n-1}{n} (1-\rho_a) \sigma_{a_{12}} \end{bmatrix}$$

$$\begin{bmatrix} b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} \frac{p-1}{p} \frac{b-1}{b} (\sigma_{e_i}^2 + \frac{\sigma_{\delta_i}^2}{n}) & \frac{p-1}{p} \frac{b-1}{b} (\sigma_{e_{12}} + \frac{\sigma_{\delta_{12}}}{n}) \\ \frac{p-1}{p} \frac{b-1}{b} (\sigma_{e_{12}} + \frac{\sigma_{\delta_{12}}}{n}) & \frac{p-1}{p} \frac{b-1}{b} (\sigma_{e_2}^2 + \frac{\sigma_{\delta_2}^2}{n}) \end{bmatrix}^{-1} \begin{bmatrix} \frac{p-1}{p} \frac{b-1}{b} (1-\rho_a) \sigma_{a_i}^2 \\ \frac{p-1}{p} \frac{b-1}{b} (1-\rho_a) \sigma_{a_{12}} \end{bmatrix}$$

$$\begin{bmatrix} b_5 \\ b_6 \end{bmatrix} = \begin{bmatrix} \frac{p-1}{p} (\sigma_{p_i}^2 + \frac{\sigma_{e_i}^2}{b} + \frac{\sigma_{d_i}^2}{nb}) & \frac{p-1}{p} (\sigma_{p_{12}} + \frac{\sigma_{e_{12}}^2}{b} + \frac{\sigma_{\delta_{12}}^2}{nb}) \\ Sim & \frac{p-1}{p} (\sigma_{p_2}^2 + \frac{\sigma_{e_2}^2}{b} + \frac{\sigma_{\delta_2}^2}{nb}) \end{bmatrix}^{-1} \begin{bmatrix} \frac{p-1}{p} \frac{1+(nb-1)}{nb} \rho_a \sigma_{a_i}^2 \\ \frac{p-1}{p} \frac{1+(nb-1)}{nb} \rho_a \sigma_{a_{12}} \end{bmatrix}$$

em que:

$\sigma_{\delta_i}^2$ e $\sigma_{\delta_{12}}^2$: variância dentro de parcelas para o caráter i e covariância dentro de parcela entre os caracteres 1 e 2, respectivamente.

$\sigma_{p_i}^2$ e $\sigma_{p_{12}}^2$: variância genética entre progênies para o caráter i e covariância genética ao nível de progênies entre os caracteres 1 e 2, respectivamente.

ρ_a : correlação genética entre os indivíduos do tipo de família considerada ($\rho_a = 0,50$, para famílias de irmãos germanos).

n , p e b : números de plantas por parcela, de progêneres e de blocos, respectivamente;

Y e X : referem-se aos caracteres 1 e 2, respectivamente.

Para o caráter 2, um índice similar a I_1 pode ser construído, somente colocando o primeiro termo dos “multiplicandos” de b_1 a b_6 como função da covariância $\sigma_{a_{12}}$ e o segundo termo dos mesmos como função da variância $\sigma_{a_2}^2$.

Como exemplo, considere os caracteres produção de frutos (y) e altura da planta em cajueiros. A genealogia e os dados encontram-se a seguir.

Indivíduos	Bloco	Pai	Mãe	Produção de Frutos (kg)	Altura
4	1	1	-	22,70	3,0
5	2	3	2	14,6	3,2
6	2	1	2	19,7	2,5
7	1	4	5	17,7	2,8
8	1	3	6	25,3	3,5

Considere ainda os seguintes componentes de variância e parâmetros genéticos:

$$h_y^2 = 0,33; \quad h_x^2 = 0,50; \quad r_{axy} = 0,20; \quad r_{xy} = 0,40;$$

$$G_o = \begin{matrix} x & \begin{bmatrix} 0,10 & 0,103 \\ 0,103 & 2,75 \end{bmatrix} \\ y & \end{matrix} \quad R_o = \begin{matrix} x & \begin{bmatrix} 0,10 & 0,28 \\ 0,28 & 5,50 \end{bmatrix} \\ y & \end{matrix}$$

$$G_o^{-1} = \begin{matrix} x & \begin{bmatrix} 10,4013 & -0,3896 \\ -0,3896 & 0,3782 \end{bmatrix} \\ y & \end{matrix} \quad R_o^{-1} = \begin{matrix} x & \begin{bmatrix} 11,6624 & -0,5937 \\ -0,5937 & 0,2120 \end{bmatrix} \\ y & \end{matrix}$$

As análises univariadas podem ser obtidas pelo mesmo sistema de equações, porém, fazendo-se:

$$G_o = \begin{bmatrix} 0,10 & 0 \\ 0 & 2,75 \end{bmatrix} \quad R_o = \begin{bmatrix} 0,10 & 0 \\ 0 & 5,50 \end{bmatrix}$$

Os resultados, tanto para a análise multivariada quanto para as análises univariadas, são apresentados na seqüência.

Efeitos	Análise Multivariada		Análise Univariada	
	Altura da planta	Produção de frutos	Altura da planta	Produção de frutos
$\hat{\beta}_1$	3,1343	22,0184	3,1373	22,0274
$\hat{\beta}_2$	2,8344	17,1696	2,8365	17,1677
\hat{a}_1	-0,1353	0,5405	-0,1117	0,4980
\hat{a}_2	0,0036	-0,0901	-0,0004	-0,0908
\hat{a}_3	0,1831	-0,2587	0,1675	-0,2056
\hat{a}_4	-0,1448	-0,0173	-0,1390	-0,0534
\hat{a}_5	0,1808	-0,9966	0,1394	-0,9386
\hat{a}_6	-0,1496	0,9574	-0,1123	0,9033
\hat{a}_7	-0,0703	-1,2506	-0,1123	-1,2623
\hat{a}_8	0,1123	0,9129	0,1393	0,9336

As acurárias (r_{aa}), tanto para a análise multivariada quanto univariada, foram obtidas a partir da expressão $r_{aa}^2 = (\sigma_a^2 - PEV) / \sigma_a^2$, onde PEV (para as equações de modelo misto em questão) é o elemento da diagonal da inversa da matriz dos coeficientes, pertencente a determinado indivíduo e caráter. Os valores de PEV e de acurácia são apresentados a seguir:

Indivíduos	Análise multivariada				Análise univariada			
	PEV		Acurácia		PEV		Acurácia	
	ALT.	PROD.	ALT.	PROD.	ALT.	PROD.	ALT.	PC
1	0,0909	2,5904	0,3011	0,2409	0,0911	2,5910	0,2977	0,2404
2	0,0973	2,7063	0,1631	0,1261	0,0974	2,7065	0,1606	0,1257
3	0,0863	2,5095	0,3702	0,2957	0,0866	2,5105	0,3659	0,2951
4	0,0773	2,3506	0,4766	0,3811	0,0778	2,3522	0,4711	0,3803
5	0,0775	2,3530	0,4739	0,3799	0,0780	2,3546	0,4686	0,3792
6	0,0814	2,4311	0,4314	0,3405	0,0819	2,4326	0,4257	0,3398
7	0,0811	2,4287	0,4344	0,3418	0,0816	2,4302	0,4285	0,3410
8	0,0759	2,3214	0,4906	0,3948	0,0764	2,3230	0,4853	0,3940

Verificam-se acurárias praticamente idênticas pelos métodos univariado e multivariado, fato que revela que, nas condições (parâmetros genéticos) estudadas, a seleção que emprega caráter auxiliar não traz vantagens.

Os valores genéticos preditos pelos métodos multivariado ou univariado podem ser utilizados na predição de valores genéticos de outros caracteres relacionados a eles, por meio da expressão (Schneeberger et al., 1992):

$\hat{g} = G_{21} G_{11}^{-1} \hat{a}$, em que G_{11} é a matriz de variância-covariância genética entre os valores genéticos em \hat{a} ; G_{21} é a matriz de covariância genética entre os valores genéticos de a e os de g e \hat{a} é o vetor de valores genéticos dos caracteres objetivos.

Ainda no contexto dos modelos multivariados, as variáveis correlacionadas podem ser transformadas em variáveis não correlacionadas, visando à realização de "c" análises univariadas em vez de uma multivariada, porém sem perda de acurácia. Dentre os procedimentos de transformação, destacam-se: (i) transformação canônica, que é adequada quando todos os caracteres são avaliados em todos os indivíduos, ou seja, quando as matrizes de incidência X e Z são as mesmas para todos os caracteres e (ii) transformação Cholesky, quando observações de alguns caracteres são perdidas e a perda é seqüencial.

7.11. BLUP e Índice de Seleção na Seleção Auxiliada por Marcadores

Uma das formas de utilização dos dados moleculares refere-se à seleção auxiliada por marcadores – MAS –, baseada em um índice de seleção formado por dados fenotípicos e dados dos marcadores (Neumann-Sorensen & Robertson, 1961; Smith, 1967; Lande & Thompson, 1990). Tal índice é da forma:

$$I_i = b_p P_i + b_M M_i, \text{ em que:}$$

P_i : informação fenotípica do indivíduo i.

$M_i = \sum_{j=1}^K \beta_j W_{ij}$: informação molecular associada ao indivíduo i.

W_{ij} : número observado de um dos alelos do loco marcador j no indivíduo i.

b_p e b_M : pesos dados às informações fenotípicas e moleculares, respectivamente.

Sendo σ_a^2 a variância aditiva dos QTL's, os pesos b_p e b_M podem ser obtidos conforme apresentado a seguir:

$$\begin{bmatrix} b_p \\ b_M \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_y^2 & \sigma_m^2 \\ \sigma_m^2 & \sigma_m^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_a^2 & \sigma_m^2 \\ \sigma_m^2 & \sigma_m^2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ ou de forma}$$

$$\text{equivalente } \begin{bmatrix} b_p \\ b_M \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \begin{bmatrix} G_{11} \\ G_{12} \end{bmatrix} = \begin{bmatrix} \sigma_y^2 & \sigma_m^2 \\ \sigma_m^2 & \sigma_m^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_a^2 \\ \sigma_m^2 \end{bmatrix}. \text{ Tomando-se } b_p = 1,$$

tem-se que:

$$b_M = \frac{F_{11} G_{12} - F_{12} G_{11}}{F_{22} G_{11} - F_{12} G_{12}} = \frac{\sigma_y^2 \sigma_m^2 - \sigma_m^2 \sigma_a^2}{\sigma_m^2 \sigma_a^2 - \sigma_m^2 \sigma_m^2} = \frac{\sigma_y^2 - \sigma_a^2}{\sigma_a^2 - \sigma_m^2} = \frac{1-h^2}{h^2-m^2} = \frac{1-h^2}{h^{2*}}$$

$g = \frac{m^2}{h^2} = \frac{\sigma_m^2}{\sigma_a^2}$: proporção da variância genética aditiva total explicada pelos marcadores.

m^2 : herdabilidade devida aos marcadores, ou seja, aos QTL's marcados;

h^2 : herdabilidade no sentido restrito ou herdabilidade poligênica devida a todos os QTL's marcados e não marcados;

σ_m^2 : variância devida aos QTL's marcados;

σ_a^2 : variância genética aditiva poligênica.

Neste caso, b_M pode ser expresso em termos da proporção g por meio de $b_M = \frac{(1/h^2 - 1)}{1-g}$.

Verifica-se por esta última expressão que o peso dado à informação molecular aumenta à medida que a herdabilidade do caráter diminui e/ou a proporção g da variância genética explicada pelos marcadores aumenta.

A eficiência relativa da MAS sobre a seleção fenotípica é dada por $E = \left[\frac{g}{h^2} + \frac{(1-g)^2}{1-h^2 g} \right]^{1/2}$. Com base nesta expressão, pode-se concluir:

- Se h^2 tende a 1, a informação molecular não acrescenta eficiência alguma.
- Se toda a variância dos QTL's for explicada pelos marcadores, $g = 1$, $b_M \rightarrow \infty$ (todo peso colocado em M_i) e a eficiência é dada por $1/h$.
- De maneira mais genérica, a eficiência da seleção baseada exclusivamente em marcadores genéticos (MBS) é dada por $[g/h^2]^{1/2}$. Assim, apenas quando a proporção g excede a h^2 , a seleção com base apenas nos marcadores é mais eficiente que a seleção fenotípica. Entretanto, com baixa h^2 , as associações com marcadores são difíceis de serem estabelecidas, de forma que, para um tamanho amostral fixo, existe uma relação direta entre a h^2 do caráter e g , ou seja, altos valores de g não ocorrem para baixos valores de h^2 (Lande & Thompson, 1990), e a MBS sempre conduzirá a menor ganho que a seleção fenotípica.

A seleção combinada clássica pode também ser estendida para a incorporação de informações moleculares:

$I = b_{p_f} P_f + b_{M_f} M_f + b_{p_w} P_w + b_{M_w} M_w$, em que:

P_f e P_w : informações fenotípicas da média de família e desvio do indivíduo em relação à média da família, respectivamente.

M_f e M_w : informações moleculares da média de família e desvio do indivíduo em relação a média da família, respectivamente.

b_{p_f} , b_{M_f} , b_{p_w} e b_{M_w} : pesos ótimos a serem dados às informações P_f , M_f , P_w e M_w , respectivamente.

Este índice equivale, para o caso balanceado, ao procedimento BLUP (que será detalhado a seguir) e reduz drasticamente a importância da informação molecular.

BLUP de QTL

Para a seleção em populações em equilíbrio, Fernando & Grossman (1989) desenvolveram um procedimento BLUP para a seleção auxiliada por marcadores – MAS –, o qual se baseia no desequilíbrio de ligação dentro de famílias. Goddard (1992b) ampliou este modelo para o caso em que o QTL encontra-se localizado entre dois marcadores (mapeamento por intervalo), obtendo um procedimento que utiliza também a matriz de parentesco do QTL associado aos marcadores. O mapeamento por intervalo permite que a informação dos valores dos segmentos cromossômicos não se perca tão rapidamente de uma geração para outra e, em teoria, esta abordagem maximiza o ganho com a MAS em qualquer programa de melhoramento.

Segundo Goddard (1992b), os procedimentos de regressão múltipla (quadrados mínimos) e máxima verossimilhança (que consideram os efeitos como fixos) para a análise de dados de marcadores genéticos e QTL's superestimam os valores genéticos preditos dos melhores indivíduos e, desta forma, o procedimento de modelos lineares mistos ou BLUP/REML apresenta grande vantagem sobre tais métodos.

Considerando um QTL localizado entre um par de marcadores, no segmento cromossômico j, tem-se que o efeito aditivo (a_i) do indivíduo i, para um caráter quantitativo é dado por:

$$a_i = \sum_j q_{ij(p)} + \sum_j q_{ij(m)} + a_i^* \quad (1), \text{ em que:}$$

$q_{ij(p)}$ e $q_{ij(m)}$: efeitos dos alelos do QTL no segmento j, alelos estes do pai e da mãe, respectivamente;

a_i^* : efeito aditivo dos QTL's não incluídos nos segmentos marcados.

O modelo BLUP usual para o valor fenotípico (Y_i) do indivíduo i equivale a:

$$Y_i = X'_i f + a_i + e_i \quad (2), \text{ em que:}$$

f e e_i : vetores de efeitos fixos e de erros aleatórios, respectivamente.

X_i : vetor de incidência para f .

Substituindo (1) em (2) tem-se:

$Y_i = X'_i f + \sum_j q_{ij(p)} + \sum_j q_{ij(m)} + a_i^* + e_i$, que expresso em notação matricial equivale a:

$$y = Xf + \sum_j Z_j q_j + Za^* + e \quad (3), \text{ em que:}$$

q_j : vetor que contém duas incógnitas para cada indivíduo em cada loco (um efeito do QTL maternal e outro do paternal).

Z_j : matriz de incidência que contém linhas com dois valores 1 e os demais 0.

Z : matriz de incidência para a^* .

Para a obtenção das equações de modelo misto do BLUP são necessárias as matrizes de covariância:

$$Var(a^*) = A \sigma_{a^*}^2$$

$$Var(q_j) = G_j = P_i \sigma_{q_i}^2 = \sigma_q^2$$

$$Var(e) = I \sigma^2, \text{ em que:}$$

A : matriz de parentesco genético aditivo.

$\sigma_{a^*}^2$: variância aditiva dos QTL's não associados aos segmentos cromossômicos marcados.

σ^2 : variância residual.

σ_q^2 : variância dos QTL's associados aos segmentos cromossômicos marcados.

P_i : matriz conhecida de parentesco do QTL marcado, conforme detalhado na seqüência.

Assim, as equações de modelo misto são dadas por:

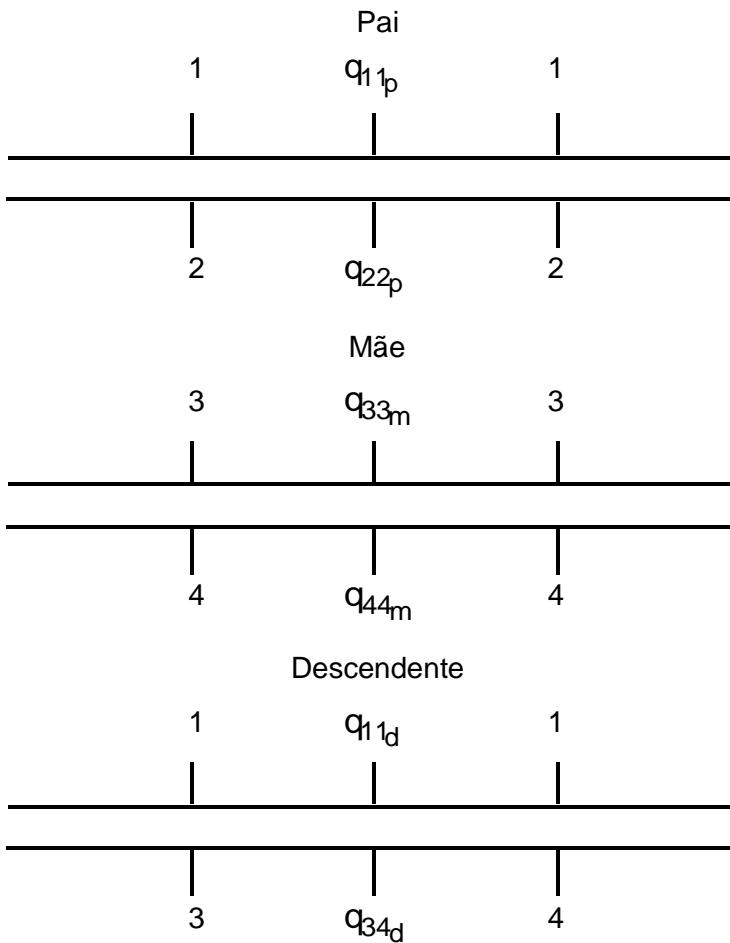
$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 & \dots \\ Z'X & Z'Z + A^{-1}\lambda & Z'Z_1 & Z'Z_2 & \dots \\ Z'_1 X & Z'_1 Z_1 & Z'_1 Z_2 & \dots \\ Z'_2 X & Z'_2 Z_1 & Z'_2 Z_2 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}^{-1} \begin{bmatrix} \hat{f} \\ \hat{a}^* \\ \hat{q}_1 \\ \hat{q}_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'_1 y \\ Z'_2 y \\ \vdots \end{bmatrix} \quad (4), \text{ em}$$

$$\text{que: } \lambda = \frac{\sigma^2}{\sigma_{a^*}^2}.$$

O número de vetores q e de matrizes Z e G é igual ao número de QTL's marcados, sendo que Z_j refere-se ao j -ésimo QTL marcado. A variância de $q(\sigma_q^2)$ é bloco diagonal, com cada bloco correspondendo a um QTL marcado

da população base em equilíbrio de ligação. Também a covariância entre a^* e q é zero se a população base encontra-se em equilíbrio de ligação.

Considere a seguinte genealogia associada a um QTL e dois marcadores:



Neste caso, considerando que não existe *crossing-over* duplo ($q_{11_p} = q_{11_d}$),

$$q_1 = \begin{bmatrix} q_{11_p} \\ q_{22_p} \\ q_{33_m} \\ q_{44_m} \\ q_{34_d} \end{bmatrix} \quad \text{e ignorando-se os efeitos fixos, o modelo linear é dado por:}$$

$$y = Z a^* + Z_1 q_1 + e$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1^* \\ a_2^* \\ a_3^* \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_{11} \\ q_{22} \\ q_{33} \\ q_{44} \\ q_{34} \end{bmatrix} + e$$

Quanto à estrutura de covariâncias:

- (i) Para os indivíduos base (genitores) tem-se $V(q_1) = V(q) = I \sigma_q^2$.
- (ii) Para os descendentes tem-se:
 - se o segmento cromossômico marcado passa do genitor para o descendente, considera-se que ele contém o mesmo alelo do QTL.
 - se um alelo do QTL pode advir de qualquer um dos genitores, supõe-se um novo alelo do QTL no descendente com efeito q_d e

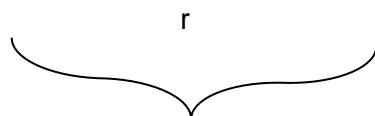
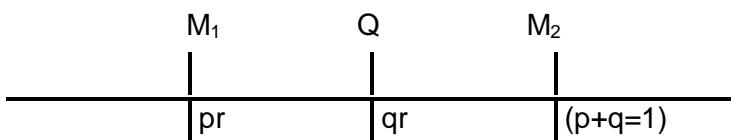
$$E(q_d) = \frac{1}{2} (q_p + q_m)$$

$$V(q_d) = \sigma_q^2$$

As considerações descritas em (i) e (ii) permitem a formação de um *pedigree* dos alelos do QTL e de uma matriz P de parentesco cuja inversa pode ser escrita diretamente de maneira similar à formação de A^{-1} .

Considerando a presente genealogia e a posição (p) do QTL dentro do intervalo marcador, a matriz P é dada por:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & q \\ 0 & 0 & 0 & 1 & p \\ 0 & 0 & q & p & 1 \end{bmatrix}, \text{ em que:}$$



Assim, considerando $p = q = \frac{1}{2}$, tem-se:

$$V \begin{bmatrix} q_{11} \\ q_{22} \\ q_{33} \\ q_{44} \\ q_{34} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 1 \end{bmatrix} \sigma_q^2 \text{ e as equações de modelo misto equivalem a:}$$

lem a:

$$\begin{bmatrix} Z'Z + A^{-1}\lambda_1 & Z'Z_1 \\ Z'_1Z & Z'_1Z_1 + P^{-1}\lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{a}^* \\ \hat{q}_1 \end{bmatrix} = \begin{bmatrix} Z'y \\ Z'_1y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma^2}{\sigma_{a^*}^2}; \quad \lambda_2 = \frac{\sigma^2}{\sigma_q^2}$$

Para o pleno uso do procedimento BLUP são necessárias as estimativas dos parâmetros p e σ_q^2 para cada loco. Os componentes de variância σ^2 , $\sigma_{a^*}^2$ e σ_q^2 podem ser estimados iterativamente nas equações de modelo misto por meio do método de máxima verossimilhança restrita – REML–, conforme Van Arendonk et al. (1994) e Grignola et al. (1996).

De posse das previsões \hat{a}^* e \hat{q}_1 , os indivíduos devem ser selecionados com base em $\hat{a}^* + Z_1 \hat{q}_1$ que, no presente exemplo, equivale a:

$\hat{a}_1^* + \hat{q}_{11} + \hat{q}_{22}$ para o pai

$\hat{a}_2^* + \hat{q}_{33} + \hat{q}_{44}$ para a mãe

$\hat{a}_3^* + \hat{q}_{11} + \hat{q}_{34}$ para o descendente.

Perspectivas do uso das informações de QTL

Uma outra abordagem molecular para a seleção de caracteres quantitativos refere-se à identificação e seleção dos próprios QTL's. Este procedimento é denominado de seleção de genes candidatos – CGS – e baseia-se em genes de grandes efeitos. Esta abordagem não requer desequilíbrio de ligação (podendo ser aplicada com pequeno tamanho de família) e tem sido descrita como o limite superior da MAS, em termos de eficiência (Muir, 1997).

A eficiência da MAS dentro de progênies depende essencialmente da fração de variância dentro da progénie que pode ser explicada pelos marcadores, e esta fração aumenta com o crescimento do tamanho da família. As estimativas da eficiência da MAS dentro de progênies sobre a seleção fenotípica são: 20% para famílias de grande ($n > 500$) tamanho; 2% a 4% para famílias de tamanho pequeno ($n \leq 30$). Assim, para a maioria dos programas de melhoramento de

espécies perenes, a eficiência da MAS é baixa, visto que pequenos tamanhos de família são empregados (Fairfull & Muir, 1997; Muir, 1997).

A seleção pelos próprios QTL's (CGS) não é vantajosa sobre a seleção fenotípica no curto e no longo prazos para caracteres com herdabilidades maiores que 1% (Muir, 1997). Com base no exposto, Fairfull & Muir (1997) concluíram que a MAS para espécies alógamas e com pequeno tamanho de famílias apresenta pouco valor prático para caracteres que podem ser medidos em todos os candidatos à seleção, portanto, é mais útil para caracteres que não podem ser medidos diretamente. Uma comparação entre a MAS e a seleção clonal é apresentada no tópico 19.7.

Quanto ao custo da MAS, deve ser levado em conta que o estabelecimento de associações marcadores-QTL não cessa após um mapa genético ter sido concluído. Para todos os sistemas marcadores, períodos limitados de uso existem e novos sistemas marcadores são necessários à medida que os alelos vão se tornando fixados em alguns locos e também devido à recombinação (Lande & Thompson, 1990).

Considerando que os custos associados à utilização da genética molecular são muito superiores aos custos associados à abordagem clássica, dentre os vários requisitos para a plena utilização das modernas técnicas moleculares, em termos comerciais, deve-se ressaltar que a alocação de recursos para estas técnicas não deve competir e comprometer a infra-estrutura do melhoramento clássico. Este fato pode ser um complicador em países tropicais (em desenvolvimento), nos quais as circunstâncias econômicas não são muito favoráveis.

7.12. Seleção Baseada em Médias Harmônicas

Os procedimentos de predição relatados nos tópicos anteriores produzem valores genéticos preditos que são funções da média aritmética das observações. Em outras palavras, o valor genético predito de um genitor é função da média aritmética das observações realizadas nos indivíduos de sua progénie. A seleção baseada em média aritmética é adequada por permitir a seleção por produtividade. Entretanto, não permite a seleção por homogeneidade. Esta se refere, no caso, à homogeneidade da progénie. A não consideração da homogeneidade dos dados é uma das propriedades indesejáveis da média aritmética, de forma que outras medidas de tendência central são recomendadas, tais quais a média harmônica (Streiner, 2000).

A predição baseada em média harmônica permite selecionar simultaneamente por produtividade e por homogeneidade, fato que é relevante no melhoramento de vários caracteres de importância agroindustrial. Apesar de problemas de ordem estatística, como a não obediência da média harmônica ao teorema central do limite, o seu uso na seleção pode ser realizado empregando-se os mesmos preditores BLUP e equações de modelo misto utilizados nos casos tradicionais de seleção baseada em média aritmética. Para isto, basta compor o

vetor de dados como a recíproca dos dados observados, ou seja, $(1/y)$. A média de $(1/y)$ fornece uma função $(1/H)$ da média harmônica (H) dos dados. Para dados com distribuição normal, maiores discrepâncias entre a média aritmética e a média harmônica ocorrem quanto maiores os desvios padrões dos dados. Quanto maior o desvio padrão, menor será a média harmônica. Dessa forma, seleção pelos maiores valores da média harmônica implica seleção também por homogeneidade. Seleções baseadas em médias harmônicas começaram a ser praticadas para melhoramento animal no Brasil (Carvalheiro et al., 2001; Muniz et al., 2001).

Outra forma de praticar a seleção por homogeneidade é por meio da consideração das variâncias dos valores genéticos preditos dos indivíduos das progênies. É importante relatar que seleção por homogeneidade pode conduzir a maior decréscimo da variabilidade genética no melhoramento de longo prazo.

7.13. Avaliação Genética em Espécies Autopoliplóides

Todas as abordagens biométricas apresentadas neste documento são válidas para espécies diplóides e alopoliplóides. O comportamento citogenético de espécies alopoliplóides é similar ao de indivíduos diplóides, ou seja, somente são formados bivalentes de cromossomos homólogos do mesmo genoma, visto que os genomas que deram origem aos alopoliplóides, em geral, diferem muito. Assim, os modelos genéticos empregados no estudo das plantas diplóides podem ser empregados diretamente em alopoliplóides, como *Coffea arabica*, *Leucaena leucocephala*, *Pennisetum purpureum* (capim-elefante).

No caso de autopoliplóides, três ou mais cromossomos homólogos existem e modificações nos modelos genéticos são necessárias. Nesta categoria de poliplóides, os autotetraplóides (quatro cromossomos homólogos) são mais comuns e relevantes, podendo-se citar a batata, o centeio, a beterraba e várias plantas forrageiras (alfafa, braquiária, capim-colonião), dentre outras. As considerações aqui realizadas se restringem aos autotetraplóides e são baseadas em Gallais (1989).

Nos autotetraplóides, o efeito genotípico de um indivíduo, ignorando a epistasia, é dado por:

$$g = a + d + t + q, \text{ em que:}$$

a: efeito genético aditivo.

d: efeito genético das interações envolvendo dois alelos ou efeito de dominância.

t: efeito genético das interações envolvendo três alelos.

q: efeito genético das interações envolvendo quatro alelos.

A variância genotípica é dada por:

$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_t^2 + \sigma_q^2, \text{ em que:}$$

σ_a^2 : variância genética aditiva.

σ_d^2 : variância das interações bialélicas ou variância de dominância.

σ_t^2 : variância das interações trialélicas.

σ_q^2 : variância das interações tetra-alélicas.

Algumas das principais covariâncias genéticas entre os parentes x e y são:

(a) Covariância genitor – descendente

$$\text{cov}_{(x,y)} = (1/2) \sigma_a^2 + (1/6) \sigma_d^2$$

(b) Covariância entre meios-irmãos

$$\text{cov}_{(x,y)} = (1/4) \sigma_a^2 + (1/36) \sigma_d^2$$

c) Covariância entre irmãos germanos

$$\text{cov}_{(x,y)} = (1/2) \sigma_a^2 + (2/9) \sigma_d^2 + (1/12) \sigma_t^2 + (1/36) \sigma_q^2$$

(d) Covariância entre ramos de um mesmo clone

$$\text{cov}_{(x,y)} = \sigma_a^2 + \sigma_d^2 + \sigma_t^2 + \sigma_q^2$$

(e) Covariância entre avô-neto

$$\text{cov}_{(x,y)} = (1/4) \sigma_a^2 + (1/36) \sigma_d^2$$

Verifica-se que todas as covariâncias diferem daquelas referentes a espécies diplóides e alloploplóides. Constatase também que, a rigor, não é possível estimar a variância genética aditiva e a herdabilidade individual no sentido restrito tendo por base apenas a avaliação de progêniess de meios-irmãos ou de genitores e filhos. Isto porque essas relações de parentesco, no caso de autotetraplóides, contemplam também frações da variância de dominância e não apenas da variância aditiva.

Para estimação dos quatro componentes de variância são necessárias ao menos quatro relações de parentesco. No entanto, pode-se assumir que σ_t^2 e σ_q^2 são de pequena magnitude e negligenciáveis (Wricke & Weber, 1986), de forma que delineamentos que geram simultaneamente progêñies de meios-irmãos e de irmãos germanos, tais quais o fatorial, o hierárquico e os dialelos, podem ser usados para estimar σ_a^2 e σ_d^2 . Como exemplo, no delineamento de cruzamento fatorial, podem ser obtidas as seguintes estimativas:

$$\hat{\sigma}_a^2 = \frac{4(p_1\hat{\sigma}_m^2 + p_2\hat{\sigma}_f^2)}{p_1 + p_2} - (2/3)\hat{\sigma}_{mf}^2$$

$$\hat{\sigma}_d^2 = 6\hat{\sigma}_{mf}^2, \text{ em que:}$$

$\hat{\sigma}_m^2$: estimativa da variância entre machos.

$\hat{\sigma}_f^2$: estimativa da variância entre fêmeas.

$\hat{\sigma}_{mf}^2$: estimativa da variância da interação entre machos e fêmeas.

p_1 : inversa da variância da estimativa de σ_m^2 .

p_2 : inversa da variância da estimativa de σ_f^2 .

Na análise via metodologia de modelos mistos, σ_a^2 é estimada diretamente pelo uso da matriz de parentesco genético aditivo e ajuste do efeito de família de irmãos germanos, cuja variância (σ_{fg}^2) estima $(2/9)\sigma_d^2$ e, portanto, σ_d^2 pode ser estimada por $\hat{\sigma}_d^2 = (9/2)\hat{\sigma}_{fg}^2$.

Os métodos para o melhoramento populacional de autotetraplóides são os mesmos adotados para as espécies diplóides. No entanto, deve-se considerar que o ganho genético imediato após cada ciclo de seleção não explora apenas a variação genética aditiva, mas também frações da variância de dominância. O ganho genético permanente, após várias recombinações, capitaliza apenas a variância genética aditiva.

Para a predição de valores genéticos pelo procedimento BLUP, deve-se considerar simultaneamente as matrizes de parentesco genético aditivo e de dominância segundo as covariâncias entre parentes apresentadas. Não é possível a predição usando apenas a matriz de parentesco genético aditivo (A), mesmo que se tenha apenas dados de meios-irmãos. Em favor de aproximações poder-se-ia ignorar a fração $(1/36)\sigma_d^2$ e considerar apenas a matriz A.

Capítulo 8

**Variáveis Binomiais e Categóricas:
Componentes de Variância
e Valores Genéticos**

8.1. Introdução

Muitos caracteres de importância econômica em várias espécies são quantitativos, mas apresentam distribuição discreta. Estas características, denominadas características de limiar, demandam considerações especiais em termos de análise quantitativa. O modelo conceitual (Lush et al., 1948; Robertson & Lerner, 1949; Dempster & Lerner, 1950) de análise de tais caracteres assume a existência de uma distribuição contínua (normal) não observável conectada à distribuição discreta observável, por meio de um ponto de truncamento ou limiar.

Os parâmetros genéticos obtidos com os dados na escala discreta podem ser utilizados visando à predição de valores genéticos e ganhos genéticos na própria escala observada. Entretanto, nesta escala existe o inconveniente de ter a dependência entre média e variância, de forma que a estimativa da herdabilidade na escala discreta é amplamente dependente da freqüência ou incidência média do fenótipo na população (Falconer, 1989). Esta estimativa de herdabilidade pode ser convertida à escala normal (Lush et al., 1948).

No contexto dos índices de seleção multicaracterísticos, envolvendo, simultaneamente, variáveis com distribuição discreta e contínua, as correlações genéticas e fenotípicas entre essas características devem também ser convertidas para a escala normal, possibilitando, assim, a construção dos índices de seleção.

No melhoramento de plantas perenes, a característica sobrevivência é muito relevante pois está diretamente relacionada com a produção por área. Esta característica é, essencialmente, adaptativa e é uma característica de limiar. A seleção desta característica e de outras variáveis binomiais necessita ser mais bem estudada, sobretudo no que se refere aos diferentes modelos de análise, à utilização da informação de família e à predição dos valores genéticos de árvores individuais.

8.2. Estimação da Herdabilidade

8.2.1. Escala binomial

Na própria escala observável, os componentes de variância e herdabilidade podem ser estimados empregando-se a tradicional (Vencovsky & Barriga, 1992) técnica da análise de variância.

Outra alternativa, que utiliza a estatística qui-quadrado (χ^2), foi apresentada por Robertson & Lerner (1949) que mostrou o estimador para a herdabilidade individual no sentido restrito:

$$h_{ab}^2 = \frac{4[\chi_{f-1}^2 - (f-1)]}{N(f-1)}, \text{ em que:}$$

χ^2_{f-1} : qui-quadrado de heterogeneidade, calculado em uma tabela de contingência 2 x f;

f e N : números de famílias e de indivíduos por família, respectivamente.

8.2.2. Escala normal

Empregando-se a transformação de probito (Bliss, 1934; Dempster & Lerner, 1950; Elston, 1977; Hill & Simith, 1977), os parâmetros estimados na escala binomial podem ser convertidos à escala normal padrão. Assim, têm-se as igualdades:

$$h_a^2 = h_{ab}^2 [p(1-p)] / z^2(1)$$

$$h_a^2 = \sigma_{a_n}^2$$

$$\sigma_{y_n}^2 = 1$$

h_a^2 e h_{ab}^2 : herdabilidades individuais no sentido restrito nas escalas normal e binomial, respectivamente.

$\sigma_{a_n}^2$ e $\sigma_{y_n}^2$: variâncias genética aditiva e fenotípica na escala normal, respectivamente.

p e z : freqüência do fenótipo desejável (ao qual é atribuído o valor 1) na população e ordenada da curva normal padrão correspondente à probabilidade de p, respectivamente.

$$z = \phi^{-1}(p); 0,1] .$$

A adequabilidade da expressão (1) foi comprovada por Van Vleck (1972) e Mc Guirk (1989), por meio de simulações.

No contexto de análise de dados categóricos, estimadores para o coeficiente de correlação foram apresentados por Tallis (1962 a, b) e Plackett (1956). Os métodos de máxima verossimilhança apresentados por Tallis são computacionalmente proibitivos, mas o estimador apresentado por Plackett, aplicável a tabelas de contingência 2 x 2 é muito simples e pode ser usado com eficiência. Este estimador, conforme será mostrado no tópico 8.3., pode ser empregado no cálculo da herdabilidade (Gianola, 1982).

8.2.3. Análise via modelos não lineares e princípios bayesianos

A análise de dados categóricos no melhoramento genético por meio do modelo de limiar pode ser realizada com base na metodologia linear, empregada sobre os dados na escala observável (binomial), de forma que a herdabilidade na escala contínua pode ser obtida usando a transformação de probito. Entretanto, conforme Gianola (1982), os parâmetros genéticos estimados por meio

desta transformação não são livres de vícios devido à não aditividade (principalmente efeitos de dominância).

Em teoria, a metodologia para caracteres contínuos não é totalmente adequada para caracteres discretos, embora propicie uma aproximação razoável. Na década passada, abordagens não lineares para a análise de dados discretos no melhoramento genético foram propostos por Gianola & Foulley (1983), Harville & Mee (1984) e Foulley et al. (1987) como forma de obtenção de estimativas mais precisas de parâmetros genéticos e preditores mais acurados de valores genéticos. Estes métodos baseiam-se, via de regra, na análise bayesiana.

Os principais métodos não lineares estudados para tal finalidade são: modelo de limiar (podendo ser assumidas diferentes distribuições para a variável na escala base ou contínua), modelo de Poisson e modelo binomial negativa. Em vários trabalhos foram comparados resultados obtidos com a metodologia de modelos lineares mistos e modelos não lineares mistos, destacando-se o de Matos et al. (1997) que compararam os modelos linear, não linear de limiar, Poisson e binomial negativa. Tais autores concluíram que os resultados das análises binomial negativa e Poisson são de difícil interpretação e de pouca utilidade prática e que o modelo não linear de limiar produz maiores estimativas de herdabilidade e, portanto, conduz a maior ganho genético que o modelo linear (a vantagem é maior quanto menor for a herdabilidade do caráter).

Com base nestes resultados, o modelo não linear de limiar tem sido bastante usado (Yazdi et al. 1999) via análise bayesiana. Entretanto, trabalho recente (Abdel-Azim & Berger, 1999), que representa o atual estado da arte da metodologia de análise de caracteres discretos, realizou uma abordagem diferente no estudo da adequabilidade do modelo não linear de limiar. Tal trabalho testou a capacidade do modelo de limiar para ajustar variáveis à escala base contínua e não observável, diferentemente dos demais que compararam soluções obtidas por diferentes modelos. Usando simulação, os autores compararam os modelos não linear de limiar e linear com transformação de dados, em termos da acurácia ou correlação entre as previsões e os valores genéticos verdadeiros (que foram simulados) na escala base contínua. Os principais resultados obtidos são a ineficiência da transformação de dados e uma maior adequabilidade do modelo não linear de limiar para caracteres com mais de duas categorias de resposta. Para variáveis binomiais (duas categorias de resposta), a determinação média do modelo não linear de limiar foi de 61% (acurácia 0,78 entre os valores genéticos preditos e os verdadeiros na escala base).

Apesar de não ser um procedimento exato, a suposição do modelo limiar e o uso da técnica de modelos não lineares para ajustar os valores observados à escala contínua não observável constituem um modelo geral mais correto e conduzem a maior herdabilidade do que a análise dos dados na própria escala binomial. A análise dos dados na própria escala binomial (como se os dados fossem normais) e posterior transformação da herdabilidade para a escala normal padrão via transformação de probito é uma boa aproximação apenas quando a incidência média do fenótipo nos diferentes níveis dos efeitos fixos (nos

diferentes blocos) forem aproximadamente iguais a p . Caso contrário, o uso da metodologia de modelos não lineares mistos é recomendado. O uso desta metodologia em associação com a técnica de modelos lineares generalizados é apresentada em detalhes no tópico 8.7.

Segundo Foulley et al. (1990), a superioridade dos modelos não lineares sobre a análise direta dos dados na escala discreta, em termos de ganho genético, é maior quando: (i) o número de categorias diminui; (ii) as incidências médias nas categorias são mais dissimilares; (iii) maior a diferença nas incidências entre os níveis dos efeitos fixos; (iv) maior o desbalanceamento; (v) maior a herdabilidade na escala base. Assim, embora o modelo não linear de limiar seja mais acurado quanto maior o número de categorias (Abdel-Azim & Berger, 1999), menos necessário é o seu uso quanto mais categorias estiverem associadas às observações (Foulley et al., 1990).

8.3. Estimação da Repetibilidade

Para caracteres binários e tendo-se duas avaliações por indivíduo, uma tabela de contingência 2×2 sumariza toda a informação amostral. Tal tabela é estruturada da seguinte forma:

Avaliação 1

		0	1	Total
Avaliação 2	0	n_{00}	N_{10}	$n_{..0}$
	1	n_{01}	N_{11}	$n_{..1}$
Total		$n_{..0}$	$N_{1..}$	$n_{..}$

8.3.1. Escala binomial

Com base nessa tabela de contingência, Lush (1956) apresentou a estatística $\rho_L = n_{11}/n_{..} - n_{01}/n_{..0}$, como sendo o estimador para a repetibilidade ou correlação entre avaliações repetidas em um mesmo indivíduo.

Para o caso da variável sobrevivência, podem ser especificadas as quantidades:

n_{11} : número de indivíduos que estavam vivos na 1^a avaliação e permaneceram vivos na 2^a avaliação, ou seja, é o próprio número de indivíduos vivos na 2^a avaliação.

$n_{..}$: soma do número de indivíduos vivos na 1^a avaliação e que morreram na 2^a mais o número de indivíduos vivos na 2^a avaliação, ou seja, é o próprio número de indivíduos vivos na 1^a avaliação.

n_{01} : número de indivíduos que estavam mortos na 1^a avaliação e que tornaram-se vivos na 2^a avaliação, ou seja, zero.

Assim, neste caso, o estimador para a repetibilidade equivale a $\rho_L = N_2 / N_1 = p_2 / p_1$, onde N_2 e N_1 são os números de indivíduos vivos na 2^a e 1^a avaliações, respectivamente, e p_2 e p_1 são as porcentagens de sobrevivência na 2^a e 1^a avaliações, respectivamente.

8.3.2. Escala normal

Considerando ainda a referida tabela de contingência, tem-se que o estimador para a correlação entre avaliações repetidas apresentada por Plackett (1965) equivale a:

$$\rho_p = \operatorname{sen} \pi/2 \left[\frac{(n_{00} n_{11})^{1/2} - (n_{01} n_{10})^{1/2}}{(n_{00} n_{11})^{1/2} + (n_{01} n_{10})^{1/2}} \right]$$

Considerando a igualdade $\operatorname{sen} \theta = -\cos(\pi/2 + \theta)$, tem-se que esta correlação é equivalente a $\rho_p = -\cos[\Psi^{1/2} \pi / (1 + \Psi^{1/2})]$, em que $\Psi = n_{00} n_{11} / n_{10} n_{01}$.

No caso da sobrevivência, n_{01} é igual a zero e, portanto, a expressão de ρ_p não se aplica. Entretanto, para uma variável como resistência a doenças, avaliada em termos de ausência e presença, o estimador ρ_p pode ser empregado para obtenção da repetibilidade, no caso em que as avaliações 1 e 2 são realizadas sobre o mesmo indivíduo. No caso em que a avaliação 1 é realizada sobre genitores e a avaliação 2 sobre a progénie, o estimador ρ_p permite obter uma estimativa da herdabilidade do caráter, na escala normal ($\rho_p = (1/2)h_a^2$).

8.3.3. Escala logística

No contexto dos modelos lineares generalizados (ver tópico 8.7.), a escala base pode ser assumida como tendo distribuição logística. Nesta situação, a repetibilidade na escala logística é dada por $\rho = \left[\frac{\sigma_a^2 + \sigma_{ep}^2}{\sigma_a^2 + \sigma_{ep}^2 + ((\pi)^2/3)} \right]$, em que σ_{ep}^2 é

a variância de ambiente permanente e $\pi = 3,14159$. A estimação da correlação intraclasse para variáveis binomiais, no contexto dos modelos lineares generalizados, é abordada com detalhes por Molenberghs et al. (1996).

8.4. Estimação das Correlações Genéticas e Fenotípicas

As correlações fenotípicas entre características de limiar (x e y) estimadas na escala binomial ($r_{(x,y)b}$), empregando-se a técnica tradicional da análise de covariância (Vencovsky & Barriga, 1992), podem ser transformadas para a escala normal padrão ($r_{(x,y)n}$), utilizando-se a seguinte relação (Olausson & Ronning, 1975):

$$r_{(X,Y) n} = r_{(X,Y) b} \left[\frac{p_x(1-p_x)}{z_x^2} \right]^{1/2} \left[\frac{p_y(1-p_y)}{z_y^2} \right]^{1/2}$$

As correlações fenotípicas envolvendo uma variável com distribuição binomial (x) e outra com distribuição normal (y) podem ser transformadas para a escala normal padrão empregando-se a relação (Olausson & Ronningen, 1975):

$$r_{(X,Y) n} = r_{(X,Y) b} \left[\frac{p_x(1-p_x)}{Z_X^2} \right]^{1/2}$$

As correlações genéticas entre dois caracteres, avaliados na escala binomial ou um caráter avaliado na escala binomial e outro na escala normal, são equivalentes nas escalas binomial e normal. O uso dos parâmetros herdabilidade e correlações, transformados para a escala normal, possibilitam a construção de índices envolvendo, simultaneamente, variáveis com distribuição normal e binomial.

8.5. Seleção e Ganho Genético na Escala Binomial

As relações de eqüivalência entre valores fenotípicos médios na escala observável (binomial) e na escala base (normal) podem ser obtidas, empregando-se propriedades da distribuição normal padrão. Assim, tem-se:

	Escala binomial	Escala normal padrão
Média geral da população	p_g	Z_g/p_g
Média da família	p_f	Z_f/p_f
Média do bloco	p_b	Z_b/p_b
Média da parcela	p_p	Z_p/p_p

Entretanto, os valores ao nível de indivíduo (0 e 1) não podem ser transformados para a escala base (normal), pois os valores 1 na escala observável eqüivalem a diferentes valores na escala base, o mesmo acontecendo com os valores de 0.

8.5.1. Seleção individual

No caso da característica sobrevivência, a seleção individual atua aleatoriamente no conjunto de indivíduos vivos da população, e a intensidade máxima de seleção eqüivale a p_g . O diferencial de seleção equivale a $(1-p_g)$ e o ganho genético é dado por $G_s = (1-p_g) h_{ab}^2$.

8.5.2. Seleção pelo índice multiefeitos

Neste caso, o índice pode ser aplicado somente sobre os indivíduos vivos e é dado por:

$$I_1 = b_1(1-p_p) + b_2(p_f - p_g) + b_3(p_p - p_f - p_b + p_g), \text{ em que:}$$

$$b_1 = \frac{(1-\rho_a) h_{ab}^2}{(1-t_1)}$$

$$b_2 = \frac{[1+(nb-1)\rho_a] h_{ab}^2}{1+(n-1) t_1 + (b-1) nt_2}$$

$$b_3 = \frac{(1-r_g) h_{ab}^2}{1+(n-1) t_1 - nt_2}, \text{ em que:}$$

$$t_2 = \rho_a h_{ab}^2;$$

$$t_1 = t_2 + c^2;$$

$$c^2 = \sigma_e^2 / \sigma_y^2.$$

ρ_a : correlação genética intraclasse entre os indivíduos da família, equivalendo a 1/4 para meios-irmãos.

σ_e^2 : variância entre parcelas.

Por este índice, os indivíduos vivos de uma mesma parcela apresentam os mesmos valores genéticos preditos.

8.5.3. Seleção de progêneres e de genitores

Para a seleção de progêneres e de genitores, podem ser construídos índices baseados nas distribuições binomial e normal:

$$I_2 = b_4 (p_f - p_g) \text{ para a binomial ou } I_3 = b_5 z_f / p_f \text{ para a normal.}$$

Para a seleção de progêneres, os estimadores são:

$$b_4 = b_2$$

$$b_5 = \frac{[1+(nb-1)\rho_a] h_a^2}{1+(n-1) t_1^* + (b-1) nt_2^*}, \text{ em que } t_1^* \text{ e } t_2^* \text{ neste caso são função de } h_a^2 \text{ e}$$

não h_{ab}^2 .

8.5.4. Seleção com medidas repetidas

Neste tipo de seleção, os estimadores são:

$$b_1 = \frac{(1-\rho_a) h_{ab}^2}{[1+(m-1) \rho - mt_1]/m}$$

$$b_2 = \frac{[1+(nb-1) \rho_a] h_{ab}^2}{[1+(m-1) \rho + (n-1) mt_1 + (b-1) nmt_2]/m}$$

$$b_3 = \frac{(1-\rho_a) h_{ab}^2}{[1+(m-1) \rho + (n-1) mt_1 - nmt_2]/m}, \text{ em que:}$$

ρ : repetibilidade.

m : número de avaliações por indivíduo.

8.6. Análise Genética da Variável Sobrevida em Espécies Florestais

Como exemplo, utilizou-se um experimento instalado no delineamento de blocos ao acaso, com 33 progênies, 6 blocos e 6 plantas por parcela. Foram avaliadas as variáveis sobrevida (em que foi atribuído o valor 1 para as plantas vivas e 0 para as mortas), altura, diâmetro (DAP) e volume cilíndrico aos 5 anos de idade.

Optou-se, neste exemplo, por considerar a abordagem tradicional baseada na análise dos dados na própria escala binomial e posterior transformação (de probito) para a escala base. O uso da própria escala binomial para variáveis com duas categorias de resposta tem revelado, em algumas situações, resultados satisfatórios em termos da coincidência entre ganhos preditos e realizados com seleção (Sohn & Goddard, 1979; White & Hodge, 1989). No exemplo em questão, a sobrevida média dos diferentes blocos não variou muito em torno da sobrevida média (Tabela 8.4), fato que corrobora a escolha do modelo de análise.

Aqui os procedimentos de predição (BLUP e índice multiefeitos) são estruturados conforme realizado para variáveis com distribuição contínua. Vale ressaltar que, no caso da sobrevida, todo experimento delineado é balanceado para esta variável, sendo que os desbalanceamentos ocorrerão apenas na comparação entre experimentos mas não dentro destes. Dessa forma, o índice multiefeitos é de grande valia na seleção para sobrevida, equivalendo ao próprio BLUP para seleção dentro de qualquer experimento (independentemente da sua taxa de sobrevida). Para a realização das análises, empregou-se o software Selegen (Resende et al., 1994a).

Na Tabela 8.1 são apresentados os resultados da análise de variância para os quatro caracteres avaliados, verificando-se a presença de diferenças significativas entre progênies.

Tabela 8.1. Quadrados médios e significâncias associadas às análises de variância para os caracteres sobrevivência (S), altura (H), diâmetro (D) e volume (V) em *Eucalyptus grandis*, bem como médias (M) e coeficientes de variação experimental (CV).

F.V.	S	H	D	V
Bloco	0,00730	49,3563	11,4294	0,04163
Progênie	0,04168*	24,4209*	11,3895*	0,01161*
Resíduo	0,01661	8,4610	4,1570	0,00506
Dentro	0,1003	41,4134	21,3861	0,0204
	M=0,8822 CV=14,61%	M=17,08 CV =17,03%	M=11,26 CV =18,11%	M=0,2389 CV =29,75%

* significativo a 1%.

Na Tabela 8.2 são apresentadas as herdabilidades para as características analisadas.

Tabela 8.2. Herdabilidades no sentido restrito ao nível de plantas (h^2), repetibilidade (ρ) e correlações intraclasse devidas ao ambiente comum da parcela (c^2) para os caracteres sobrevivência (S), altura (H), diâmetro (D) e volume (V) em *Eucalyptus grandis*.

Caráter	h^2	c^2	h_b^{2*}	ρ^{**}
S	0,4207	0,0011	0,1600	0,93
H	0,2332	0,0342	-	-
D	0,2080	0,0256	-	-
V	0,1886	0,0717	-	-

* herdabilidade na escala binomial; ** $\rho = 0,88/0,95 = 0,93$ = repetibilidade considerando a avaliação de dois anos consecutivos.

Verifica-se que a herdabilidade para a sobrevivência (na escala normal) equivale, praticamente, ao dobro das herdabilidades dos caracteres de crescimento e mais que o dobro da herdabilidade na escala binomial (Tabela 8.2). A herdabilidade de 0,42 (relativamente próxima à estimativa de 0,31 na escala normal obtida por Chambers et al., 1996) é mais realística para a população em análise, independentemente da incidência média de sobrevivência no local de experimentação. A estimativa de 0,16 é válida especificamente para este experimento com sobrevivência média de 0,8822. A repetibilidade para sobrevivência mostrou-se alta (0,93), revelando que a seleção para tal variável pode basear-se na avaliação de apenas uma idade adulta.

Outra alternativa de análise para sobrevivência poderia ser a transformação prévia das médias de parcelas para arco-seno da raiz quadrada das observações. Entretanto, os resultados práticos têm revelado que as proporções entre os componentes de variância não são afetadas pela transformação, quando a

incidência encontra-se entre 10% e 90%. Este fato, aliado ao problema da estimação da variância dentro de parcela na escala arco-seno e às concordâncias entre ganhos preditos na escala binomial e os ganhos realizados em campo, tem revelado a superioridade da análise na própria escala binomial (Sohn & Goddard, 1979; White e Hodge, 1989).

Os ganhos genéticos para sobrevivência, considerando diferentes procedimentos de seleção são apresentados na Tabela 8.3.

Tabela 8.3. Ganhos genéticos* e acurárias associadas a diferentes métodos de seleção para o caráter sobrevivência.

Método de seleção	Acurácia	Ganho genético (%)	Superioridade sobre a massal (eficiência)
Individual	0,40	2,14	1,00
Combinada	0,50	8,90	4,16
Índice Multiefeitos	0,51	9,08	4,24
Entre progênies	0,41	8,70	4,07

* considerando a seleção dos 36 melhores indivíduos.

Constata-se a grande superioridade dos métodos que utilizam a informação de família, conforme esperado teoricamente em função do baixo diferencial de seleção da seleção massal. Entretanto, para o caso da sobrevivência, apenas a seleção entre famílias também não se justifica, pois os 36 indivíduos selecionados pertencem a uma única família (que sobreviveu 100%). Tais indivíduos apresentam iguais valores genéticos preditos para sobrevivência, porém diferentes preditos para os caracteres de crescimento. Assim, os outros caracteres de interesse deveriam ser considerados simultaneamente na seleção.

Uma melhor comparação entre a seleção combinada e o índice multiefeitos pode ser feita com base nos resultados da Tabela 8.4.

Tabela 8.4. Comparação entre a seleção combinada (SC) e o índice multiefeitos (IME), considerando os indivíduos selecionados pelos diferentes métodos, bem como as médias de progênies (MP) e de blocos (MB), para a variável sobrevivência.

SC				IME							
Ordem	Bloco	Prog.	Árv.	Ordem	Bloco	Prog.	Árv.	Prog.	MP	Bloco	MB
1	5	38	1	1	3	42	1	42	1,00	3	0,85
2	5	38	4	2	3	42	2	103	1,00	2	0,87
3	5	38	5	3	3	42	3	8	0,97	5	0,88
4	5	38	6	4	5	42	4	40	0,97	6	0,89
5	3	8	1	5	3	42	5	46	0,97	4	0,89
6	3	8	2	6	3	42	6	1	0,94	1	0,90
7	3	8	4	7	3	103	1	35	0,94		
8		8	5	8	3	3	2	38	0,94		
9	3	8	6	9	3	103	3	39	0,94		
10	3	40	1	10	3	103	4	45	0,94		

Continua...

Tabela 8.4. Continuação.

SC				IME							
Ordem	Bloco	Prog.	Árv.	Ordem	Bloco	Prog.	Árv.	Prog.	MP	Bloco	MB
11	3	40	3	11	3	103	5	102	0,94		
12	3	40	4	12	3	103	6	105	0,94		
13	3	40	5	13	2	42	1	9	0,94		
14	3	40	6	14	2	42	2	4	0,92		
15	6	46	1	15	2	42	3	5	0,92		
16	6	46	2	16	2	42	4	47	0,92		
17	6	46	3	17	2	42	5	6	0,89		
18	6	46	4	18	2	42	6	10	0,89		
19	6	46	6	19	2	103	1	11	0,86		
20	1	42	1	20	2	103	2	36	0,86		
								91	0,86		
								43	0,86		
								100	0,86		
								2	0,83		

Constata-se que os 20 melhores indivíduos selecionados pela seleção combinada são diferentes dos 20 melhores selecionados pelo índice multiefeitos, ou seja, não houve coincidência alguma. O índice multiefeitos selecionou indivíduos apenas das famílias de números 42 e 103, as quais apresentaram sobrevivência de 100% e, destas famílias, os indivíduos provieram dos blocos (3 e 2) com maior mortalidade. Em resumo, o índice multiefeitos selecionou os indivíduos das melhores famílias e no ambiente de maior expressão de variabilidade (função de p ($1-p$)) e maior pressão de seleção natural.

Na seleção combinada, o componente “indivíduo dentro de parcela” no índice não foi nulo (mas o foi no índice multiefeitos, visto que o desvio $(Y_{ijk} - \bar{Y}_{ij})$ equívaleu a zero para os indivíduos selecionados) e, portanto, foram selecionados indivíduos de parcelas com mortalidade, fato que implicou a seleção de progênies intermediárias, com sobrevivência abaixo de 100%. Assim, uma vez que a seleção combinada não considera a média de bloco, esse é um procedimento inferior.

Na Tabela 8.5 são apresentados os valores genéticos aditivos preditos para volume e sobrevivência.

Tabela 8.5. Valores genéticos (média geral + efeitos aditivos) aditivos preditos para os caracteres volume (\hat{a}_{vol}), sobrevivência (\hat{a}_s) e volume por hectare (\hat{a}_{volhec}), para 20 indivíduos (superiores para volume) de *Eucalyptus grandis*.

Ordem	Bloco	Prog.	Árv.	\hat{a}_{vol} (m ³)	\hat{a}_s	Ordem	Ordem para \hat{a}_{vol}	Bloco	Prog.	Árv.	\hat{a}_{volhec} (m ³ /ha)
1	5	11	6	0,3819	0,886	1	1	5	11	6	563,94
2	4	42	6	0,3416	0,958	2	2	4	42	6	545,42
3	2	2	2	0,3320	0,872	3	4	4	8	5	521,64
4	4	8	5	0,3319	0,943	4	9	2	42	1	508,16
5	2	104	5	0,3318	0,858	5	10	4	42	4	500,71
6	5	104	3	0,3298	0,857	6	15	3	42	4	491,42
7	6	104	2	0,3244	0,856	7	17	2	42	3	485,92
8	3	104	3	0,3200	0,860	8	19	5	42	5	483,66
9	2	42	1	0,3176	0,960	9	3	2	2	2	482,51
10	4	42	4	0,3136	0,958	10	20	2	42	2	481,92
11	4	104	5	0,3135	0,855	11	5	2	104	5	474,47
12	2	104	6	0,3116	0,858	12	16	4	45	2	473,74
13	4	91	1	0,3114	0,884	13	6	5	104	3	471,06
14	2	91	1	0,3085	0,887	14	7	6	104	2	462,81
15	3	42	4	0,3065	0,962	15	13	4	91	1	458,79
16	4	45	2	0,3063	0,928	16	8	3	104	3	458,67
17	2	42	3	0,3037	0,960	17	14	2	91	1	456,07
18	4	36	4	0,3033	0,884	18	18	4	36	4	446,86
19	5	42	5	0,3026	0,959	19	11	4	104	5	446,74
20	2	42	2	0,3012	0,960	20	12	2	104	6	445,59

De acordo com essa tabela, alguns dos 20 melhores indivíduos para volume apresentam valor genético abaixo da média (0,8822) para sobrevivência. Os valores genéticos preditos dos dois caracteres mencionados foram utilizados na obtenção dos valores genéticos preditos para a variável volume por hectare, considerando o espaçamento 3 x 2 m (1.666,67 árvores por ha). Constata-se uma mudança considerável na ordem dos melhores quando se comparam as variáveis volume e volume por hectare, sobretudo em função da baixa (0,81) sobrevivência da progénie 104. Os resultados revelam a importância da consideração da variável sobrevivência nos programas de melhoramento florestal. O uso do procedimento BLUP (que penaliza os indivíduos pertencentes às progênieis que sobreviveram menos, por ocasião da predição de valores genéticos para volume por árvore) associado à obtenção dos valores genéticos preditos para sobrevivência, pode ser uma alternativa segura para contornar a questão das falhas nos experimentos de campo.

Neste trabalho, a variável volume correspondeu ao volume cilíndrico previamente calculado para cada indivíduo. Uma alternativa para predição da variável volume real consiste no uso dos valores genéticos preditos (\hat{a}) para altura e diâmetro, e posteriormente, obtenção do valor genético predito do volume real (\hat{g}) por meio da expressão: $\hat{g} = G_{21}' G_{11}^{-1} \hat{a}$, em que G_{11} refere-se à matriz de covariância genética entre os caracteres diâmetro e altura e G_{21} é relativo ao vetor de covariâncias genéticas entre volume real e altura e volume real e diâ-

metro. Estas variâncias e covariâncias genéticas podem ser obtidas a partir de herdabilidades, correlações genéticas e variâncias fenotípicas dos três caracteres mencionados. Assim, o volume real não necessita ser medido em cada indivíduo, necessitando-se apenas do conhecimento da correlação genética entre este caráter e a altura e diâmetro, para uma dada população ou família.

Outra forma para predição do valor genético para volume refere-se ao uso dos valores genéticos preditos apenas (em função da baixa precisão de campo da medição da altura) para diâmetro. Neste caso, o valor genético para volume é dado por $\hat{a}_{vol} = (COV_{a(D-V)} / \sigma_{av}^2) \hat{a}_D = (r_{a(D-V)} h_{av} / h_{aD} \sigma_{yV} / \sigma_{yD}) \hat{a}_D$, onde $r_{a(D-V)}$ é a correlação genética entre diâmetro e volume, h_{av}^2 é a herdabilidade para volume, σ_{yD} é o desvio padrão fenotípico para diâmetro e \hat{a}_D é o valor genético predito para diâmetro. Nesse exemplo, o ponderador $r_{a(D-V)} h_{av} / h_{aD} \sigma_{yV} / \sigma_{yD}$ equivale a 0,0269.

As correlações genéticas relativas aos caracteres volume (V), diâmetro (D), altura (H) e sobrevivência (S) eqüivaleram a: V-D = 0,94; V-H = 0,91; D-H = 0,99; V-S = 0,78; D-S = 0,94 e H-S = 0,97, respectivamente. Chambers et al. (1996) também obtiveram uma correlação genética positiva (0,54) entre diâmetro e sobrevivência.

Considerando a seleção multivariada (índice de seleção multicaracterístico) incluindo o volume e a sobrevivência, obtiveram-se os seguintes parâmetros: coeficiente de predição genética (herdabilidade do diferencial de seleção direto) para volume usando sobrevivência como auxiliar = 0,1914; coeficiente de predição genética para sobrevivência usando volume como auxiliar = 0,1684. Comparando-se estes valores com as herdabilidades (Tabela 8.2), têm-se as seguintes eficiências para a seleção multivariada: 1,015 (1,5% de superioridade) para volume e 1,053 (5,3% de superioridade) para sobrevivência. Assim, neste caso, existe uma leve vantagem da predição multivariada. O coeficiente de predição genética é dado por $\psi_{(x,y)} = h_{am} h_{ay}$, em que h_{ay}^2 é a herdabilidade do caráter objetivo da seleção e h_{am}^2 é a herdabilidade multivariada, a qual é dada por
$$h_{am}^2 = h_{ay}^2 \left[1 + \frac{(h_{ax} r_{a(x,y)} / h_{ay} - r_{(x,y)})^2}{r_{(x,y)}^2} \right].$$

Com base nos resultados deste exemplo, pode-se concluir que:

- (i) O índice multiefeitos ou BLUP é um método adequado para a seleção que considere variáveis com distribuição binomial a partir de experimentos balanceados, pois tende a selecionar indivíduos pertencentes às melhores famílias e que estão localizados nos blocos com maior incidência dos fenótipos indesejáveis.
- (ii) Os valores genéticos preditos para sobrevivência são muito úteis na obtenção de valores genéticos preditos para a variável volume por hectare;
- (iii) O uso do procedimento BLUP associado à seleção baseada em valores genéticos preditos para a variável volume por hectare é uma alternativa mais segura para contornar a questão das falhas nos experimentos de campo.

8.7. Estimação e Predição em Modelos Não Lineares Mistas Via Técnica de Modelos Lineares Generalizados

8.7.1. Introdução

Em teoria, variáveis categóricas ou 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 para dados dicotômicos podem ser mais apropriados.

A técnica de modelos lineares generalizados (MLG), desenvolvida por Nelder & Wedderburn (1972), permite a generalização ou flexibilização dos modelos lineares clássicos de variáveis contínuas, de forma que toda a estrutura para a estimação 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.

Variáveis binomiais relevantes na área biológica são provenientes de experimentos do tipo dose-resposta, em que os indivíduos sobrevivem, ou não, em função da dosagem do elemento adverso (produto químico, nível de estresse, concentração do inóculo etc). Neste caso, geralmente são atribuídos, por exemplo, os valores 1 para os indivíduos sobrevidentes e zero para os mortos. Entretanto, a probabilidade de ocorrer uma resposta 1 (escore 1) é nula para valores altos da dosagem e unitária para valores baixos da dosagem, de forma que tal probabilidade é uma função estritamente decrescente da dosagem. Tal função matemática é a função sigmoidal, em que os modelos são não lineares nos parâmetros, de forma que é recomendada uma transformação da curva sigmoidal em uma reta, a fim de que os procedimentos lineares possam ser aplicados na estimação dos parâmetros. Uma das transformações lineares mais recomendadas para os dados binomiais diz respeito ao uso do modelo logístico (Demétrio, 1993; Rodrigues-Zas et al., 1997).

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. Neste caso, em que o componente observational tem distribuição binomial, pode-se considerar que a distribuição da tolerância (escala contínua) em função da dosagem tem distribuição logística e a função de ligação é a própria transformação logit a qual lineariza a função sigmoidal. Pode-se, também, levar em conta que a tolerância (escala contínua) tem distribuição normal e a função de ligação que lineariza a distribuição é a função probit (Cordeiro, 1986; Demétrio, 1993).

No âmbito dos modelos mistos com efeitos fixos e aleatórios, o componente sistemático dos MLG é inerente aos próprios modelos lineares mistos clássicos, exceto o termo de erro. Neste sentido, para o caso de variáveis binomiais, a função de ligação deve ser incorporada aos estimadores dos efeitos fixos,

preditores (melhores preditores lineares não viciados – BLUP) dos efeitos aleatórios e estimadores (de máxima verossimilhança restrita – REML) dos componentes de variância dos efeitos aleatórios, conforme Gilmour et al. (1985) e Schall (1991). A função de ligação logito é a mais indicada para dados binomiais, pois constitui-se uma ligação canônica ou natural, ou seja, uma função de ligação especial para a qual existe uma estatística suficiente de mesma dimensão que β em um preditor linear $\eta = \theta = X\beta$ (Demétrio, 1993). A incorporação da função de ligação nas equações de modelos lineares mistos para a estimativa dos componentes de variância e predição de variáveis aleatórias gera a denominação de modelo não linear (Thompson, 1990), devido à relação não linear que existe entre a escala de dosagem e a probabilidade de um indivíduo pertencer a uma determinada categoria da variável binária.

Na análise de dados binomiais desbalanceados, a metodologia de modelos não lineares mistos em associação com a técnica de modelos lineares generalizados deve ser a metodologia preferida (Thompson, 1990; Knuiman & Laird, 1990; Tempelman, 1998; Gianola, 2000).

8.7.2. Variáveis binomiais e funções de transformação

Em função da correlação (pois a variância μ ($1-\mu$) é completamente determinada pela média μ) existente entre médias e variâncias de variáveis binárias (Y), o uso das técnicas padrões de modelos estatísticos lineares não é recomendado. A variável sobrevivência pode ser modelada em termos de uma escala base contínua, que é a escala da tolerância ou susceptibilidade em resposta à dosagem do ambiente adverso.

Modelada desta forma, a tolerância pode ser considerada como tendo uma distribuição logística, a qual possui forma similar a da distribuição normal.

A probabilidade μ , de uma planta sobreviver (receber o escore 1) é função decrescente da dosagem do ambiente adverso, obedecendo à função sigmóide dada por: $\mu = \frac{1}{1+e^{-\theta}}$, em que θ é a variável na escala logística.

A transformação da curva sigmóide em uma reta, visando à aplicação de procedimentos lineares é realizada por meio da transformação ou função logito dada por:

$$\begin{aligned} \eta &= \log \left[\frac{\mu}{1-\mu} \right] \\ &= \log \left[\frac{(1+e^{-\theta})^{-1}}{1-(1+e^{-\theta})^{-1}} \right] \\ &= \log \left[\frac{1}{1+e^{\theta}} \left(1 - \frac{1}{1+e^{-\theta}} \right) \right] \end{aligned}$$

$$\begin{aligned}
 &= \log \left[\frac{1}{1+e^{-\theta}} / \frac{1+e^{-\theta}-1}{1+e^{-\theta}} \right] \\
 &= \log \left[\frac{1}{1+e^{-\theta}} / \frac{e^{-\theta}}{1+e^{-\theta}} \right] \\
 &= -\log (1+e^{-\theta}) - (-\theta) - (-\log (1+e^{-\theta})) \\
 &= -\log (1+e^{-\theta}) + (\theta) + \log (1+e^{-\theta}) \\
 &= \theta,
 \end{aligned}$$

em que log é o logaritmo natural.

Assim, o logito de μ equivale a $\eta = \theta$, e um modelo linear pode ser imposto a η .

Substituindo-se os valores obtidos de θ na função sigmóide $[\mu = e^\theta / (1+e^\theta)]$, produz-se μ no intervalo entre 0 e 1.

8.7.3. Modelos lineares generalizados

Estes modelos foram desenvolvidos para análise de dados associados a distribuições pertencentes à família exponencial com um parâmetro. Nelder & Wedderburn (1972) introduziram a idéia de modelos lineares generalizados visando permitir maior flexibilidade de análise. Tal idéia relaxa a suposição de que Y segue distribuição normal e permite que esta siga qualquer distribuição que pertença à família exponencial na forma canônica. As generalizações ocorrem em duas direções: (i) permitem que a esperança μ , de Y , seja uma função monotonicamente diferenciável do preditor linear $\eta = \sum x_i \beta_i$ de forma que $\mu = f(\eta) = f(\sum x_i \beta_i)$; (ii) ou, por inversão, $g(\mu) = \eta$, em que g é a função de ligação, a qual liga a média ao preditor linear.

Para dados binomiais, $0 < \mu < 1$, funções de ligação tal qual a logito são utilizadas para satisfazer esta restrição natural. As transformações são importantes para: (i) estender a amplitude da variável analisada de $(0,1)$ para a reta real; (ii) fazer a variância constante através da amplitude dos efeitos fixos (na escala de tolerância). A função de ligação descreve, então, a relação existente entre o preditor linear (η) e o valor esperado μ de Y . No modelo linear clássico, tem-se $\eta = \mu$ que é chamada ligação de identidade, e esta ligação é adequada no sentido em que ambos η e μ podem assumir valores na reta real (McCullagh & Nelder, 1989).

O modelo linear generalizado é, então, definido por (Demétrio, 1993):

- (a) Um componente aleatório representado pelas variáveis aleatórias independentes Y_1, Y_2, \dots, Y_n provenientes de uma mesma distribuição que faz parte da família exponencial na forma canônica.
- (b) Um componente sistemático ou determinístico linear nos parâmetros, chamado preditor linear,

$\eta = X\beta$, por exemplo, em que:

$$\eta' = (\eta_1, \eta_2, \dots, \eta_n);$$

$$\beta' = (\beta_1, \beta_2, \dots, \beta_p): \text{vetor de parâmetros};$$

$X' = (X_1, X_2, \dots, X_n)$, sendo X_i 'vetor de variáveis explanatórias;

$$X_i' = (X_{i1}, X_{i2}, \dots, X_{ip}).$$

- (c) uma função de ligação $g(\cdot)$ tal que:

$$\eta_i = g(\mu_i), \text{ em que}$$

$$\mu_i = E(Y_i)$$

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.

Segundo Cordeiro (1986), a palavra generalizado significa uma distribuição mais ampla que a normal para a variável resposta e uma função não-linear conectando a média desta variável com a parte determinística do modelo.

8.7.4. Estimação e predição em modelos não lineares

As distribuições a serem assumidas para a escala de tolerância e correspondentes funções de ligação devem ser capazes de transformar o intervalo $(0,1)$ em $(-\infty, \infty)$. Neste sentido, as distribuições logística, normal padrão e Gumbel (ou distribuição de valor extremo) para a tolerância e suas correspondentes funções de ligação denominadas logito, probito e complemento log-log são apropriadas para o modelo binomial. Destas, as distribuições logística e normal padrão são simétricas e a Gumbel não o é. A função de ligação complemento log-log ($\log[-\log(1-\mu)]$) é mais indicada para valores de μ próximos de zero.

Existe pouca diferença entre as distribuições normal e logística para a tolerância ou escala contínua (Cordeiro, 1986). Entretanto, a função de ligação logito é a mais indicada para dados binomiais, pois se refere a uma ligação canônica ($\eta = \theta$) ou natural, tendo uma interpretação mais simples. A função de ligação complemento log-log, é a mais indicada para valores de μ próximos de

zero, casos em que o complemento log-log e a logito são equivalentes. Outras funções usuais de ligação para a binomial são a arcsen $\sqrt{\mu}$ e a identidade. A ligação identidade é um caso especial da ligação potência $\eta = \mu^\lambda$, em que λ é um número real. Pela identidade tem-se $\lambda = 1$ e portanto $\eta = \theta = \mu$. No modelo linear clássico (normal), a ligação é a identidade, no sentido de que os valores esperados dos dados e os preditores lineares podem ter qualquer valor real. Entretanto, para dados binomiais, a ligação identidade não é atrativa, pois não produz a transformação adequada, do intervalo (0,1) para os reais.

Para variáveis binárias, pode-se, ocasionalmente, considerar o modelo de Poisson, conforme Foulley et al. (1987). Na Tabela 8.6 são apresentadas algumas distribuições e funções de transformação associadas a dados binários.

Tabela 8.6. Distribuições para o componente aleatório (variável observada) e escala de tolerância, funções de ligação, bem como variâncias ambientais e herdabilidades na escala de tolerância, para variáveis binárias (binomiais ou Poisson) e normais.

Componen- te aleatório	Escala de tolerância	Funções de ligação	Variância ambiental	Herdabilidade
Binomial	Logística	Logito: $\eta = \log [\mu/(1-\mu)]$	$\pi^2/3$	$h^2 = 4\sigma_p^2/(\sigma_p^2 + \pi^2/3)$
	Normal Padrão	Probito: $\eta = \Phi^{-1}(\mu)$	1	$h^2 = 4\sigma_p^2/(\sigma_p^2 + 1)$
	Gumbel	Comp log-log: $\eta = \log[-\log(1-\mu)]$	$\pi^2/6$	$h^2 = 4\sigma_p^2/(\sigma_p^2 + \pi^2/6)$
	Binomial	Identidade: $\eta = \mu$	$\mu(1-\mu)$	$h^2 = 4\sigma_p^2/[(\sigma_p^2 + [\mu(1-\mu)])]$
Poisson	-	Log (μ)	μ^{-1}	$h^2 = 4\sigma_p^2/(\sigma_p^2 + \mu^{-1})$
Normal	-	Identidade: $\eta = \mu$	σ_e^2	$h^2 = 4\sigma_p^2/(\sigma_p^2 + \sigma_e^2)$

σ_p^2 : variância genética entre progêneres de meios-irmãos.

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 (função logística) a média μ_i do dado binário y_i ;

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

β e a : parâmetros de locação na escala logito (escala de tolerância).

X e Z : matrizes de incidência para β e a , respectivamente.

Não existe termo de erro em η , diferentemente dos modelos lineares. A variância residual, ou seja, a variância das observações, dado η_i (ou μ_i), advém da amostragem binomial, ou seja:

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

8.7.5. Modelos não lineares fixos

No caso da variável de trabalho definida pelo modelo $y = X\beta + Za + (y - \mu)$ ($g'(\mu)$ que com erro normal e função de ligação identidade equivale ao modelo linear clássico $y = X\beta + Za + e$) de efeitos fixos, onde os a 's são tratados como parâmetros fixos, se cada y_i é uma binomial independente ($1, \mu_i$), Nelder & Wedderburn (1972) mostraram que as soluções para as estimativas de máxima verossimilhança de (β, a) são obtidas resolvendo-se iterativamente:

$$\begin{bmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'W y^* \\ Z'W y^* \end{bmatrix}, \text{ em que:}$$

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

$$\mu = E(Y);$$

U : uma matriz diagonal com elementos $u_i = (\partial \eta_i / \partial \mu_i)$;

W : uma matriz diagonal com elementos $w_i = (\partial \mu_i / \partial \eta_i)^2 / [\mu_i(1-\mu_i)]$.

Especificamente para a ligação logito, têm-se:

$$\mu_i = \frac{e^{\eta_i}}{(1+e^{\eta_i})}; \quad w_i = \mu_i(1-\mu_i); \quad u_i = 1/w_i.$$

Assim, $y_i^* = \eta_i + \frac{y_i - \mu_i}{\mu_i(1-\mu_i)}$. Genericamente, esta expressão é equivalente a:

$g(y) = g(\mu) + (y - \mu) g'(\mu) = y^*$, em que $g(\mu)$ é a função de ligação:
 $g(\mu) = \eta = \log\left(\frac{\mu}{1-\mu}\right)$.

O procedimento iterativo é também denominado de mínimos quadrados ponderados iterativos (IWLS). Um algoritmo empregando o método numérico de Newton-Raphson é dado por:

- (1) Forneça os valores iniciais para $\hat{\mu}_i^{(0)}$ e $\hat{\eta}_i^{(0)} = g(\mu_i^{(0)})$;
- (2) Dado $\hat{\mu}_i^{(t)}$ e $\hat{\eta}_i^{(t)}$, forme a variável dependente ajustada, $\hat{y}_i^{*(t)} = \hat{\eta}_i^{(t)} + (y_i - \hat{\mu}_i^{(t)})g'(\hat{\mu}_i^{(t)})$, e também $\hat{\mu}_i^{(t)} = \mu_i^{(t)}(1-\mu_i^{(t)})$;
- (3) Reunindo β e a em um único vetor ϕ e X e Z em uma única matriz de incidência Q , obtenha $\hat{\phi}^{(t+1)}$ por $\hat{\phi}^{(t+1)} = (Q'\hat{W}^{(t)}Q)^{-1}Q'\hat{W}^{(t)}y^{*(t)}$, em que $\hat{W}^{(t)} = \text{diag}(w_i^{(t)})$. Defina $\hat{\eta}^{(t+1)} = Q\hat{\phi}^{(t+1)}$ e $\hat{\mu}^{(t+1)} = g^{-1}(\hat{\eta}^{(t+1)})$;

- (4) Repita os passos 2 e 3 até que se atinja a convergência segundo o critério desejado.

Conforme Firth (1991), os valores óbvios são $\hat{\mu}_i^{(o)} = y_i$, $\hat{\eta}_i^{(o)} = g(\mu_i^{(o)})$, em que y_i é a proporção observada.

Empregando-se o software MATLAB, tem-se a partir da iteração 0:

$$\hat{\phi}^{(1)} = \text{inv}(Q^*(\text{diag}(w_o)) * Q) * (Q \text{diag}(w_o))^* y_o^*;$$

$$\eta = Q^* \hat{\phi}^{(1)};$$

$$\mu = (e.^{\eta}) ./ (1 + e.^{\eta});$$

$$j = \text{one}\{N, 1\}; \quad (N \text{ é o número de elementos do vetor } \hat{\phi})$$

$$w = \mu * (j - \mu);$$

$$y = \eta + (y - \mu) ./ w.$$

8.7.6. Modelos não lineares mistos

Em $\eta = X\beta + Za$, define-se:

β e a : vetores de efeitos fixos e aleatórios, respectivamente;

X e Z : matrizes de incidência para β e a , respectivamente.

Todos os efeitos são assumidos como não correlacionados e os efeitos aleatórios são assumidos como tendo esperança zero e matriz de covariância G .

A variável de trabalho pode ser definida pelo modelo $y = \mu + (y - \mu)g'(\mu)$, em que $(y - \mu)g'(\mu)$ 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 $y = X\beta + Za + e$, em que $E(y) = X\beta$, $Cov(a) = G$, $Cov(e) = R = I\sigma_e^2$ e, consequentemente, $Cov(y) = ZGZ + R = ZGZ + I\sigma_e^2$.

Para este modelo linear misto tradicional, têm-se as seguintes equações de modelo misto para a estimativa de β e predição de a :

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

Na situação em que os efeitos aleatórios são não correlacionados, os estimadores para os componentes de variância pelo método da máxima verossimilhança (ML) são dados por (Fellner, 1986; 1987):

$$\hat{\sigma}_a^2 = \frac{\hat{a}'\hat{a}}{q - \text{tr } C_{22}^{-1}/\sigma_a^2}; \quad \hat{\sigma}_e^2 = \frac{(y - X\hat{\beta} - Z\hat{a})'(y - X\hat{\beta} - Z\hat{a})}{N - r(x) - q + \text{tr } C_{22}^{-1}/\sigma_a^2} \sigma_e^2$$

Pelo método de máxima verossimilhança restrita (REML), os estimadores são dados por:

$$\hat{\sigma}_a^2 = \frac{\hat{a}'\hat{a}}{q - \text{tr } C^{22}/\sigma_a^2}; \quad \hat{\sigma}_e^2 = \frac{(y - X\hat{\beta} - Z\hat{a})'(y - X\hat{\beta} - Z\hat{a})}{N - r(x) - q + \text{tr } C^{22}/\sigma_a^2} \sigma_e^2, \text{ em que:}$$

C_{22} : partição da matriz dos coeficientes das equações de modelo misto, referentes aos efeitos aleatórios;

C^{22} : partição da inversa da matriz dos coeficientes das equações de modelo misto, referentes aos efeitos aleatórios;

q : número de níveis do efeito aleatório a;

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

tr : operador traço matricial;

σ_e^2 e σ_a^2 : valores correntes ou atuais (obtidos na iteração anterior) de $\hat{\sigma}_e^2$ e $\hat{\sigma}_a^2$, respectivamente.

Estes estimadores são idênticos aos apresentados por Dempster et al. (1977) e Harville (1977), referentes ao algoritmo EM (Expectation – Maximization).

Para as situações em que Y segue uma distribuição binomial, tem-se o caso da estimação 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 $y^* = X\beta + Za + (y - \mu)g'(\mu)$, em que:

$$E(y^*) = X\beta, \quad \text{Cov}(a) = G, \quad \text{Cov}[(y - \mu)g'(\mu)] = W^{-1}\sigma_e^2 \quad \text{e} \quad \text{Cov}(y^*) = ZGZ + W^{-1}\sigma_e^2.$$

O modelo $y^* = X\beta + Za + (y - \mu)g'(\mu)$ tem a mesma estrutura da primeira e segunda ordem que o modelo $y = X\beta + Za + e$, 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 $Cov[(y-\mu)g'(\mu)]=W^{-1}\sigma_e^2$.

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

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

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*);

β_L e a_L = efeitos fixos e aleatórios na escala de tolerância.

Os estimadores dos componentes de variância pelo método ML são dados por:

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

Pelo método REML, os estimadores são dados por:

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

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:

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

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 μ 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 μ (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

$$\mu, \sigma_{e_L}^2 \text{ e } \sigma_{a_L}^2.$$

- d) Obtenção de $\hat{\sigma}_{e_L}^2$ e $\hat{\sigma}_{a_L}^2$, iterativamente e, após a convergência, proceder à obtenção atualizada de $\hat{\beta}_L$ e \hat{a}_L .
- e) Obtenção de $\hat{\eta} = \hat{\theta} = X\hat{\beta}_L + Z\hat{a}_L$.
- f) Obtenção de novo valor predito de μ , 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)).
- g) Atualização de S^{-1} via $S^{-1} = \hat{\mu}_l(1-\hat{\mu}_l) \frac{1}{\hat{\sigma}_{e_L}^2}$ e de y^* via
- $$y^* = \log \left(\frac{\hat{\mu}_l}{1-\hat{\mu}_l} \right) + \frac{y - \hat{\mu}_l}{\hat{\mu}_l(1-\hat{\mu}_l)} = \hat{\theta} + \frac{y - \hat{\mu}_l}{\hat{\mu}_l(1-\hat{\mu}_l)}.$$
- h) 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). Algoritmo hierárquico similar (mas não envolvendo variáveis binomiais) para modelos não lineares foi apresentado por Gregorie & Schabenberger (1996).

Os algoritmos apresentados são do tipo EM. 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). Para GLMM's, os procedimentos de estimativa e predição de Schall (1991), Breslow & Clayton (1993) e Engel & Keen (1994) são equivalentes (Keen & Engel, 1997). Sob suposições de normalidade, para componentes de variância fixos, o procedimento IRREML é equivalente ao procedimento bayesiano MAP (máximo a posteriori), de forma que o IRREML fornece uma alternativa, não bayesiana, de derivação do MAP (Engel & Keen, 1996).

O uso de modelos não lineares é especialmente indicado quando os indivíduos a serem avaliados pertencem a diferentes níveis dos efeitos fixos, com diferentes valores de incidência para a variável em análise. Neste caso, a obtenção do coeficiente de herdabilidade na escala base contínua, por meio da transformação de probito é inapropriada, pois a transformação é função da incidência, a qual difere para os indivíduos dos diferentes níveis dos efeitos fixos. Uma vez que as diferenças entre os níveis dos efeitos fixos correspondem a mudanças na escala base, uma função para ligação das mudanças nas duas escalas necessita ser incorporada nas equações de modelo misto. Para cômputo de y^* e S^{-1} , diferentes valores de μ_i devem ser computados para os diferentes níveis dos efeitos fixos.

8.7.7. Aplicação a dados experimentais

Foram considerados dados de dois experimentos conduzidos pela Empresa Brasileira de Pesquisa Agropecuária – Embrapa. Um deles refere-se a um teste de progêneres de *Eucalyptus grandis*, instalado no delineamento de blocos ao acaso com 33 progêneres de meios-irmãos, 6 repetições e 6 plantas por parcela. O outro, a um teste de progêneres de erva-mate (*Ilex paraguariensis*) instalado no delineamento de blocos ao acaso com 20 progêneres de meios-irmãos, 8 repetições e 6 plantas por parcela. Em ambos os experimentos, foi avaliada a variável sobrevivência das plantas, nas idades de dois anos para o experimento de erva-mate e cinco anos para o experimento com eucalipto.

Na avaliação da sobrevivência foram atribuídos os escores 1 para as plantas vivas e 0 para as mortas. O modelo misto associado às observações dos dois experimentos, quando o erro é normal e a função de ligação identidade, equivale a:

$$y = X\beta + Za + e, \text{ em que:}$$

y : vetor de dados.

β : vetor dos efeitos fixos de blocos.

a : vetor paramétrico dos efeitos aleatórios de plantas.

e : vetor de erros aleatórios.

A estrutura de médias e variâncias do modelo equivale a:

$$E \begin{bmatrix} y \\ a \\ e \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} y \\ a \\ e \end{bmatrix} = \begin{bmatrix} V & ZA\sigma_a^2 & I\sigma_e^2 \\ A\sigma_a^2 Z' & A\sigma_a^2 & 0 \\ I\sigma_e^2 & 0 & I\sigma_e^2 \end{bmatrix}, \text{ em que:}$$

A = matriz de correlação genética aditiva entre as plantas.

I = matriz identidade.

Este modelo é parcimonioso, pois não contém o efeito de parcela, o qual mostrou-se não significativo pelo teste da razão de verossimilhança aplicado, considerando os máximos do logaritmo das funções de verossimilhança associados à estimativa de efeitos fixos e componentes de variância e predição dos efeitos aleatórios, de modelos incluindo e não incluindo o efeito de parcela.

As análises foram realizadas empregando-se os estimadores, preditores e algoritmos apresentados nos itens 8.7.5 (considerando todos os efeitos fixos) e 8.7.6 (considerando o modelo misto). Utilizou-se o software ASREML (Gilmour et al., 2000), o qual emprega o algoritmo AI-REML (*Average Information REML*), desenvolvido por Gilmour et al. (1995) e Johnson & Thompson (1995).

8.7.7.1. Experimento 1 – Eucalipto

Na Tabela 8.7 são apresentados os resultados dos efeitos fixos de blocos, para o experimento com eucalipto, considerando a função de ligação logito, com os resultados das médias de blocos transformados para a escala original.

Tabela 8.7. Estimativas para as médias de blocos, para a variável sobrevivência no experimento de *Eucalyptus grandis*.

Blocos	Média ± desvio padrão
1	0,8990 ± 0,0256
2	0,8687 ± 0,0256
3	0,8535 ± 0,0256
4	0,8939 ± 0,0256
5	0,8737 ± 0,0256
6	0,8889 ± 0,0256

Significância dos efeitos de blocos: $F = 0,57$, o qual não é significativo.

Verifica-se que os blocos apresentaram médias de magnitudes próximas, as quais são estatisticamente iguais (Tabela 8.7).

Na Tabela 8.8 são apresentados os resultados dos componentes de variância associados aos efeitos aleatórios de plantas e residuais, considerando a técnica de modelos lineares generalizados mistos, usando diferentes funções de ligação.

Tabela 8.8. Estimativas (pelo método REML) dos componentes de variância referentes aos efeitos aleatórios de plantas (σ_a^2) e do erro (σ_e^2), bem como do coeficiente de determinação (herdabilidade individual no sentido restrito) dos efeitos aleatórios de plantas [$h_a^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$], associados a diferentes modelos de análise de acordo com a função de ligação empregada. Experimento de *Eucalyptus grandis*.

Distribuição de tolerância	Função de ligação	($\hat{\sigma}_a^2$)	($\hat{\sigma}_e^2$)	(\hat{h}_a^2)
Logística	Logito	1,60922	2,08298	0,43584
Normal Padrão	Probito	0,45854	0,65609	0,41138
Valor Extremo (Gumbel)	Complemento log-log	0,28659	1,42999	0,16696
Distribuição da escala original	Função de ligação	($\hat{\sigma}_a^2$)	($\hat{\sigma}_e^2$)	(\hat{h}_a^2)
Binomial	Identidade	0,01881	0,08701	0,1776
Normal	Identidade	0,01834	0,08795	0,1725

Verifica-se que a análise dos dados na própria escala observada ou original, assumindo as distribuições normal ou binomial (ambas com função de ligação identidade) conduziram a estimativas da herdabilidade próximas (na faixa de 0,17 a 0,18) (Tabela 8.8). Considerando as distribuições logística, normal padrão e de Gumbel para a variável na escala transformada (usando as funções

de ligação logito, probito e complemento log-log, respectivamente), as estimativas da herdabilidade obtidas foram de aproximadamente 0,44; 0,41 e 0,17, respectivamente (Tabela 8.8).

Constata-se que o uso de funções de ligação logito e probito são altamente favoráveis, pois conduziram a maior precisão (determinação) dos efeitos aleatórios de plantas, conforme revelado pelas estimativas de h_a^2 . Este fator é muito significativo tendo em vista que o objetivo do experimento é a seleção das plantas com os maiores efeitos aleatórios preditos. A transformação usando o complemento log-log mostrou-se inadequada, pois conduziu a h_a^2 de mesma magnitude que a análise dos dados na escala observada.

Os resultados obtidos pelo uso das transformações logito e probito foram similares, conforme foi verificado em outros experimentos (Villalobos & Garrick, 1999). Entretanto, pode-se optar pelas previsões obtidas pelo uso da função logito, tendo em vista a maior estimativa de h^2 obtida (0,44 contra 0,41) e também o fato de a ligação ser canônica e, portanto, a mais recomendada para dados binomiais.

Quando a sobrevivência média é a mesma em todos os níveis dos efeitos fixos (blocos), a h_a^2 na escala normal padrão (h_n^2) pode ser obtida a partir da própria h_a^2 na escala observada (h_{ab}^2) pela expressão (Thompson, 1990): $h_n^2 = h_{ab}^2 [\mu (1-\mu)]/z^2$, em que μ é a sobrevivência média e z é a ordenada da curva normal padrão no ponto igual a μ . Neste experimento, $h_{ab}^2 = 0,1725$ e $\mu = 0,8796$ e portanto $z = 0,20121$ e $h_n^2 = 0,4512$. Este valor difere (superestima) um pouco do valor mais realístico (0,41138), o qual foi obtido considerando que as sobrevivências diferem nos blocos. Assim, apesar de as médias dos blocos não terem apresentado diferenças estatisticamente significativas, o uso da técnica de modelos lineares generalizados mostrou-se ainda vantajoso.

Os resultados das previsões das médias de tratamentos (progêniens ou genitores) são apresentados na Tabela 8.9, considerando três abordagens: como efeitos fixos na escala logística, como efeitos aleatórios na escala observada e como efeitos aleatórios na escala logística.

Considerando as previsões na escala logística como as mais realísticas e adequadas, verifica-se que a previsão direta com dados na escala observável conduziu a resultados bastante similares aos obtidos pelo procedimento ótimo (Tabela 8.9). Este fato já era esperado tendo em vista a não significância do efeito de blocos e o balanceamento do experimento. Conforme Foulley et al. (1990), a superioridade dos preditores não lineares aumenta com o aumento do desbalanceamento dos experimentos e com o aumento da discrepância entre as incidências nos diferentes níveis dos efeitos fixos.

Tabela 8.9. Médias de tratamentos estimadas (EB) e preditas (PB) na escala observada (binomial) e preditas na escala logística (PL), no experimento de *Eucalyptus grandis*.

Tratamento (genitor)	EB	PB	PL
1	0,9444	0,9197	0,9202
2	0,8333	0,8510	0,8554
3	0,7778	0,8166	0,8187
4	0,9167	0,9025	0,9054
5	0,9167	0,9025	0,9054
6	0,8889	0,8853	0,8896
7	0,9722	0,9369	0,9339
8	0,9444	0,9197	0,9202
9	0,8889	0,8853	0,8896
10	0,8611	0,8681	0,8729
11	0,9444	0,9197	0,9202
12	0,8611	0,8681	0,8729
13	0,8333	0,8510	0,8554
14	0,9444	0,9197	0,9202
15	0,9444	0,9197	0,9202
16	0,9722	0,9369	0,9339
17	0,8056	0,8338	0,8373
18	1,0000	0,9541	0,9463
19	0,8611	0,8681	0,8729
20	0,6944	0,7650	0,7599
21	0,9444	0,9197	0,9202
22	0,9722	0,9369	0,9339
23	0,9167	0,9025	0,9054
24	0,8611	0,8681	0,8729
25	0,7778	0,8166	0,8187
26	0,7500	0,7994	0,7995
27	0,6667	0,7478	0,7395
28	0,8611	0,8681	0,8729
29	0,7778	0,8166	0,8187
30	0,9444	0,9197	0,9202
31	1,0000	0,9541	0,9463
32	0,8056	0,8338	0,8373
33	0,9444	0,9197	0,9202

8.7.7.2. Experimento 2 – Erva-mate

Na Tabela 8.10 são apresentados os resultados relativos aos efeitos fixos de blocos, para o experimento com erva-mate, considerando a função de ligação logito, com os resultados das médias de blocos transformados para a escala original.

Tabela 8.10. Estimativas para as médias de blocos, para a variável sobrevivência no experimento de erva-mate (*Ilex paraguariensis*).

Blocos	Média + desvio padrão
1	0,8250 ± 0,0386
2	0,8833 ± 0,0386
3	0,8083 ± 0,0386
4	0,7333 ± 0,0386
5	0,8833 ± 0,0386
6	0,8583 ± 0,0386
7	0,8833 ± 0,0386
8	0,6750 ± 0,0386

Significância dos efeitos de blocos: $F = 16,22$, o qual é altamente significativo.

Os blocos apresentaram médias estatisticamente diferentes entre si (Tabela 8.10).

Na Tabela 8.11 são apresentados os resultados relativos aos componentes de variância associados aos efeitos aleatórios de plantas e do erro, considerando a técnica de modelos lineares generalizados mistos, usando diferentes funções de ligação.

Conforme verificado também para o experimento com *Eucalyptus grandis*, a análise direta dos dados na própria escala observada (considerando distribuição normal e função de ligação identidade) e o uso das funções de ligação identidade e complemento log-log não conduziram a resultados satisfatórios em termos de h_a^2 , produzindo as menores magnitudes das estimativas deste parâmetro. O uso das funções de ligação logito e probito levou a resultados similares e de maior magnitude para \hat{h}_a^2 , mostrando-se mais adequado (Tabela 8.11).

Tabela 8.11. Estimativas (pelo método REML) dos componentes de variância dos efeitos aleatórios de plantas (σ_a^2) e do erro (σ_e^2), bem como do coeficiente de determinação (herdabilidade individual no sentido restrito) dos efeitos aleatórios de plantas [$h_a^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$], associados a diferentes modelos de análise de acordo com a função de ligação empregada. Experimento de erva-mate (*Ilex paraguariensis*).

Distribuição de tolerância	Função de ligação	($\hat{\sigma}_a^2$)	($\hat{\sigma}_e^2$)	(\hat{h}_a^2)
Logística	Logito	1,22625	2,37018	0,3410
Normal padrão	Probito	0,37500	0,71875	0,3429
Valor extremo (Gumbel)	Complemento log-log	0,25092	1,45674	0,1469
Distribuição da escala original	Função de ligação	($\hat{\sigma}_a^2$)	($\hat{\sigma}_e^2$)	(\hat{h}_a^2)
Binomial	Identidade	0,02305	0,12569	0,1550
Normal	Identidade	0,02728	0,11734	0,1886

A obtenção de h_a^2 na escala normal padrão, considerando a sobrevida média $\mu = 0,81873$, pela expressão $h_n^2 = h_{ab}^2 [\mu(1-\mu)]/z^2$, conduziu a $\hat{h}_n^2 = 0,4026$. Esse valor é bastante superestimado em relação ao valor mais realístico de 0,3429, o qual foi obtido considerando que a sobrevida difere ao longo dos blocos. Estes resultados, em comparação com aqueles obtidos para o experimento com eucalipto, permitem confirmar que o uso da abordagem de modelos lineares generalizados mistos é tanto mais recomendada quanto maiores forem as diferenças entre as incidências da variável através dos níveis dos efeitos fixos.

Os resultados relativos às previsões dos efeitos de tratamentos (progêneres ou genitores) são apresentados na Tabela 8.12.

São observadas previsões similares pelos procedimentos PB (escala observável) e PL (escala logística) também para o experimento com erva-mate (Tabela 8.12). A consideração dos efeitos de tratamentos como fixos conduziu a superestimativas das médias de tratamentos para ambos os experimentos (Tabelas 8.9 e 8.12).

Tabela 8.12. Médias de tratamentos estimadas (EB) e preditas (PB) na escala observada (binomial) e preditas na escala logística (PL), no experimento de erva-mate (*Ilex paraguariensis*).

Tratamento (Genitor)	EB	PB	PL
1	0,9375	0,9023	0,9044
2	0,8333	0,8290	0,8415
3	0,8542	0,8436	0,8550
4	0,9375	0,9023	0,9044
5	0,8333	0,8290	0,8415
6	0,7917	0,7996	0,8135
7	0,8125	0,8143	0,8277
8	0,6458	0,6970	0,7061
9	0,9375	0,9023	0,9044
10	0,6667	0,7117	0,7222
11	0,8958	0,8730	0,8807
12	0,7083	0,7410	0,7536
13	0,8750	0,8583	0,8681
14	0,8333	0,8290	0,8415
15	0,8958	0,8730	0,8807
16	0,8333	0,8290	0,8415
17	0,8125	0,8143	0,8277
18	0,8125	0,8143	0,8277
19	0,5833	0,6530	0,6571
20	0,8750	0,8583	0,8681

De maneira genérica, pode-se inferir que o uso da metodologia de modelos não lineares mistos associado à técnica de modelos lineares generalizados é um procedimento conceitualmente mais adequado. Na pior das hipóteses, tal procedimento conduz a resultados similares aos obtidos pela abordagem linear normal. Na análise simultânea de vários experimentos desbalanceados, o uso dos modelos não lineares é fortemente recomendado.

Se o conjunto de dados apresenta estrutura não linear (e provavelmente não normal), a aplicação direta dos modelos lineares torna-se imprópria. Neste caso, têm-se duas opções: (i) a transformação dos dados de forma que estes sejam adequados (ou se moldem) ao modelo linear; (ii) o ajustamento de modelos adequados aos dados. A abordagem (ii) é preferida e caracteriza a técnica de modelos lineares generalizados.

8.7.7.3. Parâmetro de dispersão e ajuste da distribuição discreta aos dados

Além de sua utilidade na estimação, o princípio da verossimilhança também permite comparar a adequabilidade de vários modelos, desde que tenham uma estrutura hierárquica ou aninhada. Dados dois modelos U e V com máximos das funções de verossimilhança (restrita ou não) $L(U)$ e $L(V)$ e correspondentes números de parâmetros n_u e n_v , tem-se que, menos duas vezes o logarítmico da razão de verossimilhança, $D = -2 \log(L(U))/L(V)$ possui, aproximadamente, distribuição proporcional a uma χ^2 com $n_v - n_u$ graus de liberdade (assumindo U como hierárquico ou um caso especial de V). Quando o modelo V é saturado e equívale a uma exata reprodução dos dados, D recebe o nome especial de *deviance* do modelo U (Crosbie & Hinch, 1985). Sob normalidade de erros, a *deviance* equívale a soma de quadrados dos resíduos, a qual segue uma distribuição múltipla de uma χ^2 .

Para dados com distribuição normal, uma análise de variância baseada nas somas dos quadrados para uma seqüência de modelos pode ser utilizada para investigar o ajuste de vários modelos e dos efeitos das variáveis explicativas, fatores e suas interações. Para os modelos lineares generalizados, um procedimento análogo (Nelder & Wedderburn, 1972), baseado nos valores de D , e não em soma de quadrados dos resíduos, tem sido utilizado para construção de uma análise de *deviance*.

A função *deviance* definida anteriormente tem algumas propriedades simples que indicam a sua utilidade como uma medida da qualidade do ajustamento. Se um modelo se ajusta perfeitamente aos dados, isto é, $\hat{\mu} = y$, a *deviance* é zero; caso contrário, é positiva. Além disso, uma vez que a maximização da verossimilhança de qualquer modelo é equivalente à minimização da *deviance* (Firth, 1991), o procedimento de máxima verossimilhança propicia o melhor ajuste pelo critério da *deviance*.

Pode-se notar (Firth, 1991) que a *deviance* pode ser considerada por si só como a diferença $D(y, \hat{\mu}) - D(y, y)$ entre as *deviances* para o modelo ajusta-

do e para um conceitual modelo saturado com $\hat{\mu} = y$. Dessa forma, a *deviance* fornece a qualidade do ajustamento quando se substitui os valores y pelos valores ajustados $\hat{\mu}$.

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 ϕ . Para dados binários, ϕ 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 ϕ 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 (μ) 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 ϕ (Firth, 1991; Paul & Islam, 1995).

Segundo Lindsey (1997), a *deviance* média para o modelo maximal pode propiciar uma estimativa para o parâmetro de dispersão ϕ . Esta *deviance* média é denominada fator de heterogeneidade ou fator de heterogeneidade de variância (Lindsey, 1997; Gilmour et al., 2000). Tal fator de heterogeneidade é então calculado por $\hat{\phi} = D/df$, em que D é a *deviance* como uma medida da falta de ajuste da distribuição binomial e df refere-se aos números de graus de liberdade associados ao cálculo de D . Dessa forma, $\hat{\phi}$ fornece uma indicação da adequabilidade da distribuição discreta (binomial no caso) para descrever os dados.

Nos exemplos apresentados, as estimativas de ϕ equivaleram a, aproximadamente, 0,90 para ambos os experimentos. Tendo em vista que o domínio de ϕ é de 0 a ∞ e que o valor que indica um ajuste perfeito é 1, pode-se concluir que o valor 0,90 obtido corrobora a adequabilidade da distribuição binomial aos dados.

Com base nos resultados apresentados, pode-se concluir para os experimentos em questão: os estimadores, preditores e algoritmos apresentados permitem uma acurada predição de efeitos aleatórios e uma precisa estimação de componentes de variância associados a variáveis binomiais, por meio do uso da técnica de modelos lineares generalizados mistos; as funções de ligação logito e probito mostraram-se adequadas na análise dos dados de sobrevivência de plantas de espécies perenes; as funções de ligação complemento log-log e identidade são inadequadas aos dados binomiais analisados; a função de ligação logito é a mais adequada, pois conduziu a coeficientes de determinação (h^2) dos efeitos aleatórios similares aos obtidos pela função probito e refere-se a base canônica ou natural; as estimativas do parâmetro de dispersão ou fator de heterogeneidade apresentaram magnitudes da ordem de 0,90, fato que indica uma subdispersão muito pequena e que corrobora a adequabilidade da distri-

buição binomial aos dados de sobrevivência; quando os experimentos são balanceados e as incidências (sobrevivências) nos diferentes níveis dos efeitos fixos são aproximadamente iguais, não existe grande prejuízo em usar uma análise direta dos dados na escala observada; na análise de dados desbalanceados provenientes de experimentos diversos, com diferentes sobrevivências, recomenda-se o uso de modelos não lineares e dos algoritmos apresentados.

Capítulo 9

**Estimação de Componentes
de Variância via REML**

A predição de valores genéticos e os métodos de seleção dependem, essencialmente, de estimativas fidedignas de componentes de variância, os quais são assumidos como conhecidos nos procedimentos BP, BLP e BLUP, mas na prática devem ser estimados com a maior precisão e acurácia possíveis.

9.1. Propriedades dos Estimadores

Em inferência estatística paramétrica, os estimadores devem apresentar as seguintes propriedades desejáveis:

- a) Não vício, tal que a esperança matemática do estimador seja o próprio parâmetro.
- b) Consistência, tal que, com o aumento do tamanho da amostra, a esperança do estimador converja para o parâmetro e a variância do estimador, para zero (associado ao conceito de convergência em probabilidade): refere-se ao aumento da acurácia de uma estimativa com o aumento do tamanho da amostra.
- c) Eficiência, tal que o estimador apresente variância mínima.
- d) Suficiência, tal que o estimador condense o máximo possível a informação contida na amostra e não seja função (dependente) do parâmetro.
- e) Completitude que está ligada à unicidade do estimador.
- f) Invariância à translação, tal que a estimação dos componentes de variância não seja afetada por mudanças nos efeitos fixos.

Em inferência estatística freqüentista, os principais métodos de estimação são: método dos momentos (quadrados mínimos), máxima verossimilhança, máxima verossimilhança restrita e método \mathcal{R} .

9.2. Principais Métodos de Estimação em Inferência Estatística Freqüentista

9.2.1. Quadrados mínimos (LS)

Os procedimentos usuais de estimação de componentes de variância são explícitos e baseiam-se no método de quadrados mínimos, tais como os métodos I, II e III descritos por Henderson (1953). Estes procedimentos permitem trabalhar com um certo nível de desbalanceamento e se reduzem à usual análise de variância quando os dados são balanceados. Os métodos de Henderson baseiam-se na suposição de amostragem aleatória, a qual, dificilmente, se verifica em dados obtidos em programas de melhoramento de espécies perenes. Assim, as estimativas obtidas são freqüentemente tendenciosas devido à seleção empregada.

O método LS, na situação de dados desbalanceados (situação comum em melhoramento de espécies perenes) apresenta a propriedade de não vício, mas as demais propriedades desejáveis não são asseguradas (Searle, 1971). Outra

desvantagem do método é a possibilidade de obtenção de estimativas negativas de componentes de variância. Assim, em casos de grandes desbalanceamentos de dados, a utilização do método não é recomendada.

9.2.2. Máxima verossimilhança (ML)

O método da máxima verossimilhança baseia-se na obtenção do ponto de máximo de uma função de verossimilhança (que é a função densidade de probabilidade conjunta dos pontos amostrais). E este máximo é obtido por derivação da função de verossimilhança (L) em relação ao parâmetro de interesse. O método ML foi desenvolvido por Fisher (1922), mas somente após cerca de 45 anos, Hartley & Rao (1967) apresentaram a especificação matricial de um modelo misto, fato que permitiu a derivação de equações ML para várias classes de modelos. Segundo Searle (1991), os trabalhos de Henderson (1953) tiveram grande impacto no desenvolvimento dos métodos de estimação de componentes de variância a partir de dados desbalanceados, estimulando principalmente os trabalhos de Hartley e Rao.

Em situações de dados desbalanceados, os estimadores ML apresentam as seguintes propriedades desejáveis: suficiência, consistência, eficiência e invariância a translação. Outra vantagem do ML é a geração de estimativas não negativas dos componentes de variância. Por outro lado, os estimadores ML são viciados em decorrência da perda de graus de liberdade devida à estimação dos efeitos fixos (Shaw, 1987). A estimação ML de componentes de variância considera que todos os efeitos fixos são conhecidos sem erro.

Embora viciado, o procedimento ML é computacionalmente mais simples que o método REML (descrito a seguir) e, em determinadas situações, apresenta eficiência satisfatória. O vício pode ser considerável se o número de equações independentes (posto de X), para os efeitos fixos for relativamente grande em relação ao número (N) de observações. Quando o posto de X é pequeno em relação a N , os métodos ML e REML conduzem a resultados similares, conforme verificado por Resende et al. (1996b).

9.2.3. Máxima verossimilhança restrita (REML)

O método REML foi desenvolvido por Patterson & Thompson (1971), que apresentaram uma correção ao ML, eliminando o seu vício. No método REML, somente a porção da verossimilhança que é invariante aos efeitos fixos é maximizada. Assim, o REML mantém as demais propriedades do ML, é não viciado e permite também a imposição de restrições de não negatividade. Dessa forma, o REML é o procedimento ideal de estimação de componentes de variância com dados desbalanceados. A idéia inicial de maximizar a parte da função de verossimilhança que é invariante aos parâmetros de locação do modelo, isto é, aos efeitos fixos, foi de Thompson Jr. (1962), mas isto, em termos de uma base mais geral, é devido a Patterson e Thompson. Também Robertson (1962), con-

siderando a estimativa de componentes de variância a partir de dados desbalanceados em modelos de classificação simples (única), derivou pesos ótimos para as médias de família, os quais são, efetivamente, aqueles usados na estimativa REML. O procedimento REML é também denominado de máxima verossimilhança marginal e, na Europa, por máxima verossimilhança residual.

Os métodos ML e REML requerem solução iterativa, pela não linearidade das equações, fato que dificulta a derivação de estimadores explícitos. Assim, os componentes de variância iteragem nas equações de modelo misto do BLUP até a convergência para um valor adequado. Na maioria das vezes, a estimativa dos componentes de variância e a predição de valores genéticos são realizados, simultaneamente, pelo procedimento REML/BLUP. Neste caso, o procedimento BLUP é denominado BLUP empírico.

Os métodos de máxima verossimilhança permitem, sob certas condições, levar em conta os efeitos da seleção. É necessário que todas as informações que tenham contribuído para a seleção sejam incluídas na análise (bem como o parentesco entre os indivíduos), exceto se essas informações não forem correlacionadas com o caráter sob análise. Mesmo se essas condições forem apenas parcialmente obedecidas, o método fornece estimativas menos tendenciosas que aquelas obtidas pelo método III, de Henderson, ou outros métodos baseados em análise de variância (Meyer, 1989a). De maneira geral, argumentos teóricos e evidências indicam que inferências pontuais realizadas a partir de funções de verossimilhança não são afetadas por algumas formas de seleção (Gianola et al., 1989). Essa propriedade fez com que o método REML se tornasse padrão para a estimativa de componentes de variância em programas de melhoramento genético.

O método REML elimina o vício devido a mudanças nas freqüências alélicas pela seleção, pelo uso da matriz de parentesco completa (A). Assim, torna-se possível a obtenção de componentes de variância para uma população base não selecionada e a predição de valores genéticos de indivíduos de quaisquer gerações é realizada com precisão. O uso da matriz A completa leva em consideração as alterações na variância genética devidas à endogamia e ao desequilíbrio de ligação, as quais resultam da seleção (Kennedy & Sorensen, 1988) e, considera também, a tendência genética ou ganho genético realizado.

Os métodos REML e ML exigem normalidade para que os estimadores tenham propriedades desejáveis. Entretanto, tais estimadores podem ser robustos aos desvios da normalidade, gerando estimativas razoáveis mesmo quando a forma da distribuição não é especificada (Harville, 1977). O método LS (análise de variância) apresenta a propriedade de não vício independentemente da normalidade ou não dos dados (Lynch & Walsh, 1997).

Para dados balanceados, o estimador REML (negligenciando a suposição de normalidade e a restrição de estimativas REML ao espaço paramétrico) é idêntico ao estimador LS (análise de variância), o qual apresenta as propriedades desejáveis BUE (melhor estimador não viciado) e BQUE (melhor estimador quadrático não viciado) sob normalidade (Searle et al., 1992).

9.2.4. Método \mathfrak{R}

Recentemente, o método \mathfrak{R} , baseado no coeficiente de regressão linear (\mathfrak{R}) entre previsões atuais (\hat{a}) e prévias (\hat{a}_p) de valores genéticos individuais, foi proposto como uma alternativa computacionalmente mais simples de estimação de componentes de variância (Reverter et al., 1994). No caso, as previsões prévias são obtidas a partir de subamostras (aproximadamente 50%) do conjunto total de dados. O coeficiente de regressão (\mathfrak{R}) é computado por $\mathfrak{R} = (\hat{a}' A^{-1} \hat{a}_p) / (\hat{a}_p' A^{-1} \hat{a}_p)$, em que A é a matriz de parentesco entre os indivíduos.

A esperança de \mathfrak{R} equivale a 1 e o uso de uma proporção (função da herdabilidade) de componentes de variância inadequada causa um desvio positivo ou negativo de 1. Se o valor de \mathfrak{R} computado for superior a 1, a razão entre os componentes de variância empregados foi subestimada e se \mathfrak{R} for inferior a 1, a referida razão entre os componentes de variância foi superestimada. Assim, várias iterações são realizadas até a obtenção de \mathfrak{R} igual a, aproximadamente, 1.

As esperanças das estimativas dos componentes de variância obtidas pelo procedimento \mathfrak{R} são equivalentes às estimativas REML. Dessa forma, o procedimento \mathfrak{R} tem sido utilizado, com freqüência no melhoramento animal, sobre tudo em modelos que incluem os efeitos de dominância em que o número de equações é muito grande. O método tem a vantagem de que apenas as soluções das equações de modelo misto são necessárias, sendo que a inversa da matriz dos coeficientes não é utilizada.

As propriedades estatísticas das estimativas obtidas pelo método \mathfrak{R} não são ainda muito claras (Hofer, 1998). Reverter (1998) comparou o BLUP empírico usando estimativas de componentes de variâncias obtidas por REML e pelo método \mathfrak{R} , e não verificou diferença entre os dois procedimentos. Segundo Misztal (1997), o método \mathfrak{R} apresenta as seguintes propriedades desejáveis: (i) produz estimativas dentro do espaço paramétrico; (ii) é resistente a vícios causados por cruzamentos preferenciais visto que a matriz A é considerada; (iii) é tão resistente quanto o REML para diferentes formas de vícios devidos à seleção.

Por outro lado, Schenkel (1999) relata que o método \mathfrak{R} pode não ser tão robusto aos efeitos da seleção quanto o método REML, mesmo quando se usa a matriz de parentesco completa. Isto porque Van Der Werf & De Boer (1990) mostraram que as alterações na variância genética causada pelo desequilíbrio em fase gamética são consideradas completamente apenas quando todas as observações utilizadas nas decisões de seleção forem usadas na estimação de parâmetros genéticos. Uma vez que o método \mathfrak{R} baseia-se em subamostragens do conjunto completo de dados, as consequências do desequilíbrio em fase gamética não seriam consideradas por este método.

9.3. Algoritmos para a Obtenção do Máximo da Função de Verossimilhança

A implementação computacional da metodologia de modelos mistos baseia-se fortemente em métodos numéricos, notadamente, em álgebra linear numérica, visando à obtenção iterativa das soluções das equações de modelo misto (obtenção do BLUP) e, cálculo numérico para a maximização/ minimização de funções de várias variáveis, visando à obtenção das estimativas REML.

Vários algoritmos computacionais para a obtenção de componentes de variância por ML e REML têm sido desenvolvidos tais como o MS (Method of Scoring de Fisher), o EM (Expectation-Maximization, de Dempster et al., 1977), o DF-REML (Derivative-free Restricted Maximum Likelihood, de Graser et al., 1987 e o AI-REML (Average Information-REML de Johnson & Thompson, 1995). Dentro destes, os mais usados são o EM, o DF-REML, e o AI-REML. O algoritmo EM é muito estável, numericamente, apresentando convergência mesmo que os valores iniciais não tenham sido totalmente adequados. Entretanto, uma inconveniência do algoritmo EM é a lentidão para as estimativas próximas ao limite do espaço paramétrico (por exemplo, quando uma variância tende a zero). Se valores iniciais positivos forem utilizados, a convergência para valores não negativos é garantida (Harville, 1977).

O algoritmo EM atua por meio da obtenção da esperança (por integração) e maximização (derivação) da função de verossimilhança dos dados, sucessivamente. Nos modelos de plantas individuais, em que, freqüentemente, a ordem das equações de modelo misto excedem o número de observações, a obtenção de estimativas por meio de primeira derivada pelo método EM requer a inversão da matriz dos coeficientes das equações de modelo misto, aumentando muito o esforço computacional. Segundo Lynch & Walsh (1997), os métodos de Newton-Raphson e de Fisher apresentam convergência quadrática, ao passo que o algoritmo EM apresenta convergência linear, sendo, portanto, mais lento.

Para contornar esta questão, Graser et al. (1987) propuseram um algoritmo para obtenção do ponto de máximo do logaritmo da função de verossimilhança por meio de sucessivas avaliações da função, partindo de valores atribuídos aos componentes de variância. Assim, o máximo relativo aos componentes de variância é encontrado por um processo de procura direta, sem requerer a inversão da matriz dos coeficientes. Por não envolver a derivação da função densidade de probabilidade, em relação aos componentes de variância, para o estabelecimento do sistema de equações a ser utilizado no processo iterativo, o algoritmo foi denominado DFREML. O algoritmo DF requer menos tempo de processamento por iteração do que o EM, porém exige maior número de iterações para atingir a convergência.

Recentemente, o algoritmo denominado AI-REML (Average Information-REML) foi desenvolvido e tem sido muito utilizado (Gilmour et al., 1995; Johnson & Thompson, 1995; Meyer, 1997). Segundo Johnson & Thompson (1995), o algoritmo DF apresenta pobres propriedades numéricas, e as soluções apresentam baixa acurácia nos dígitos significativos o que constitui um problema quando vários componentes de variância são estimados. Assim, o algoritmo AI é mui-

to competitivo em relação aos algoritmos DF e EM e converge em menos de 10 iterações, embora o tempo por iteração seja 2 a 3 vezes maior que aquele requerido pelo algoritmo DF. Mesmo assim, o tempo total para convergência é muito menor. Segundo os últimos autores, para modelos complexos, o algoritmo AI é cinco vezes mais rápido que o DF e três vezes mais rápido que o algoritmo EM.

Os algoritmos para obtenção de estimativas REML podem ser agrupados de acordo com a ordem das derivadas usadas. Assim, têm-se: (i) não derivativo (DF-REML); (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 hessiano (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.

Uma comparação empírica dos algoritmos REML foi realizada por Hofer (1998) e os resultados são apresentados na Tabela 9.1.

Tabela 9.1. Resultados da comparação empírica de algoritmos para a obtenção de estimativas REML em termos do número (I) de iterações (ou avaliações de funções, para o algoritmo DF) e tempo total (T) de computação até a convergência, em horas.

NMME	NVC	DF		EM		AI	
		I	T	I	T	I	T
4.895	3	26	0.01	24	0.05	-	-
9.780	9	238	0.31	33	0.26	-	-
14.685	18	583	1.77	45	1.02	-	-
6.192	9	699	1.27	-	-	-	-
10.230	12	1.236	2.33	-	-	-	-
14.274	18	4.751	11.10	-	-	-	-
5.731	5	169	0.34	109	1.14	6	0.07
8.765	6	927	70.60	23	4.97	7	1.86
5.073	2	39	0.02	-	-	5	0.02
10.146	6	472	0.52	-	-	9	0.09
233.796	55	37.021	2083	-	-	-	-

NMME = número de equações de modelo misto; NVC = número de componentes de variância.

Os algoritmos DF ganharam popularidade devido às suas flexibilidades quanto aos modelos (Meyer, 1989; 1991) e por causa da disponibilidade de softwares. Entretanto, as dificuldades de convergência em modelos mais complexos geraram um novo interesse em algoritmos baseados em primeira e segunda derivadas da função de verossimilhança, como o AI. Pela Tabela 9.1, torna-se claro que os algoritmos DF são vantajosos computacionalmente somente quando o número de parâmetros a serem estimados é pequeno. Os algoritmos EM são mais eficientes que o DF quando o número de parâmetros é grande. Nesta mesma situação, o algoritmo AI supera o EM.

9.4. Programas Computacionais

Programas e algoritmos computacionais para a estimação REML de componentes de variância e predição de valores genéticos, simultaneamente, têm sido desenvolvidos com o emprego dos algoritmos EM, DF e AI.

Os primeiros programas computacionais desenvolvidos empregaram o algoritmo EM (Danell, 1987; Robinson, 1987). Posteriormente, Meyer (1988) desenvolveu o programa DFREML e Boldman et al. (1995) desenvolveram o programa MTDFREML (Multiple Trait-DFREML) ambos empregando o algoritmo DF. Outros programas desenvolvidos foram o PEST/VCE (Groeneveld et al., 1990) e JAA/MTC (Misztal, 1994). Recentemente, foram desenvolvidos os softwares ASREML (Gilmour & Thompson, 1998), o qual utiliza o algoritmo AI, e MTGSAM (Van Tassel & Van Vleck, 1995), que fornece estimativas obtidas por métodos bayesianos. O aplicativo MTDFREML emprega o método numérico denominado simplex (para a minimização de uma função de várias variáveis) ao passo que o DFREML emprega os métodos de Powell e o simplex pelo algoritmo DF e, também o algoritmo AI.

Os problemas computacionais associados à estimação REML referem-se à inversão de matrizes e obtenção de primeiras e segundas derivadas necessárias à avaliação da maximização de funções, fatores esses que demandam grandes recursos computacionais em termos de tempo de processamento e memória. Este fato tem levado os pesquisadores a utilizarem diferentes técnicas de análise numérica visando diminuir a necessidade de recursos computacionais associada ao uso do procedimento REML.

Dentre os softwares mencionados, os mais utilizados no melhoramento de plantas perenes no Brasil são o ASREML, o DFREML e o MTDFREML. Estes aplicativos 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 documento são apresentados aspectos da utilização dos softwares DFREML (Meyer, 1998) e ASREML (Gilmour et al. 2000). Os programas do DFREML (Versão 3.0β) 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 destina-se à recodificação dos efeitos fixos e aleatórios na ordem de processamento e à construção da matriz de parentesco; DFUNI, dedicado a análises univariadas, permitindo o ajuste de vários efeitos aleatórios não correlacionados, adicionais aos efeitos aditivos; DXMUX, para análise multivariada, que permite também o ajuste de diferentes modelos para os caracteres individuais, os quais podem diferir tanto nos efeitos fixos quanto aleatórios. Este programa possibilita o ajuste de, no máximo, dois efeitos aleatórios adicionais não correlacionados; DXMRR destinado à 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. Detalhes do software ASREML são apresentados no tópico referente à estimação de componentes de variância para variáveis binomiais e categóricas.

9.5. Estimadores ML e REML para Modelos com um Efeito Aleatório, Empregando o Algoritmo EM

Para um modelo misto com um só efeito aleatório, além do erro, tal como um modelo para a predição dos efeitos genéticos aditivos individuais, a partir da avaliação de progêniés de meios-irmãos no delineamento de blocos ao acaso com uma planta por parcela, tem-se:

Modelo:

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

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

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

Equações de Modelo Misto:

$$\begin{bmatrix} X'X & X'Z \\ ZX & ZZ + A^{-1}(\sigma_e^2 / \sigma_a^2) \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \end{bmatrix}$$

Estimador ML

Segundo o modelo apresentado e o algoritmo EM, os estimadores para obtenção de σ_a^2 e σ_e^2 (Schaeffer, 1976) são:

$$\hat{\sigma}_e^2 = [y'y - \hat{b}'X'y - \hat{a}'Zy]/N$$

$$\hat{\sigma}_a^2 = [\hat{a}'A^{-1}\hat{a} + \sigma_e^2 \operatorname{tr}[Z'Z + A^{-1}(\sigma_e^2/\sigma_a^2)]^{-1}]/N_a, \text{ em que:}$$

N_a : número de elementos aleatórios (indivíduos).

A : matriz de parentesco genético aditivo.

tr : operador traço matricial, dado pela soma dos elementos da diagonal da matriz.

N : número total de dados.

Assim, a partir de um valor inicial de σ_e^2/σ_a^2 , estima-se \hat{b} e \hat{a} pelas equações de modelo misto, os quais são usados para obtenção de novas estimativas de σ_e^2 e σ_a^2 . Repete-se este procedimento até a convergência para valores estáveis. A velocidade de convergência é tanto maior quanto mais acertada (próxima da verdadeira) for a informação inicial utilizada.

Estimador REML

Para o mesmo modelo, o algoritmo EM conduz aos seguintes estimadores de σ_a^2 e σ_e^2 (Harville, 1977; Meyer, 1989):

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

$$\hat{\sigma}_a^2 = [\hat{a}'A^{-1}\hat{a} + \sigma_e^2 \operatorname{tr} C^{22}] / N_a, \text{ em que:}$$

$r(X)$: posto ou número de colunas linearmente independentes de X .

$$C^{22} \text{ é da forma } \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ ZX & ZZ + A^{-1}(\sigma_e^2/\sigma_a^2) \end{bmatrix}^{-1}.$$

Os componentes de variância iteragem nas equações de modelo misto até a convergência, conforme descrito para o método ML.

9.6. Estimadores DF-REML

9.6.1. Estimadores DF-REML para modelos com um efeito aleatório

O algoritmo DF-REML para modelos com apenas um fator aleatório além do resíduo, proposto por Graser et al. (1987), determina o ponto de máximo da função L por meio de sucessivas avaliações dessa função para valores de q (razão entre os componentes de variância genética aditiva e residual). Assim, esse algoritmo não envolve a derivação da função L em relação aos componentes de variância, para a determinação dos estimadores dos componentes de variância. A seguir, será descrito o procedimento DFREML, para o modelo de plantas individuais com apenas um efeito aleatório.

Para um valor a priori de $q = \sigma_a^2 / \sigma_e^2$, pode-se estimar:

$$\hat{\sigma}_e^2 = \frac{y' Py}{N - r(x)} = \frac{y'(y - \hat{x}b - \hat{Z}a)}{N - r(x)}, \text{ em que:}$$

$P = V^1 - V^1 X (X' V^1 X)^{-1} X' V^1$: matriz de projeção ortogonal da parte aleatória das observações (y) no espaço coluna da matriz X .

$$V = ZAZ \hat{\sigma}_a^2 + I \hat{\sigma}_e^2 = Var(y)$$

Para estimar $\hat{q} = \hat{\sigma}_a^2 / \hat{\sigma}_e^2$, são calculadas apenas as partes relevantes do máximo (q fixo) de L para vários valores de q , ou seja, L é maximizado com respeito apenas ao parâmetro q . Para avaliar a máxima verossimilhança para qualquer q , deve-se, também, avaliar $\hat{\sigma}_e^2$ e $\hat{\sigma}_a^2 = q\hat{\sigma}_e^2$. Neste caso, a função L a ser avaliada equivale a:

$$L = -\frac{1}{2} [(N - r(x) - N_a) \log_e \hat{\sigma}_e^2 + \log_e |C| + N_a \log_e \hat{\sigma}_a^2 + y' Py / \hat{\sigma}_e^2], \text{ em que:}$$

$|C|$: determinante de uma submatriz não singular de C com posto máximo, em que C é a matriz dos coeficientes das equações de modelo misto.

N_a : número de níveis no efeito aleatório a .

Assim, uma vez que $\hat{\sigma}_e^2$ tenha sido calculado usando um valor a priori de q , a estimativa REML de \hat{q} pode ser obtida diretamente da equação acima. Para avaliação de L é necessária a obtenção apenas de $y' Py$ e $\log_e |C|$, uma vez que os demais termos são fixados.

Como a inversão de V (em P) demanda grande esforço computacional, Graser et al. (1987) propuseram uma estratégia computacional que baseia-se na absorção de C em $y'y$ de forma a se obter $y' [V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}] y = y' Py$. A absorção de C em $y'y$ não requer inversão, ou seja, é realizada pelo processo de eliminação de Gauss, absorvendo uma linha por vez, conforme Smith & Graser (1986). Este método apresenta a vantagem adicional de se obter $\log_e |C|$ pelo somatório do logaritmo dos elementos da diagonal C , ou seja, por meio dos pivôs não zero obtidos durante o processo de absorção via eliminação de Gauss.

Na prática três valores de q , em torno de seu valor esperado, devem ser utilizados. Uma função quadrática (equação de regressão, usando os três valores de q) pode ser calculada visando descrever L como função de q , e o ponto de máximo (valor de q que maximiza L) da equação quadrática fornece um novo valor de q . A seguir, deve-se reavaliar L com este novo valor de q e recalcular a equação quadrática usando este novo valor de q e mais dois adjacentes dentre os três anteriores. O procedimento deve ser repetido até a convergência para o valor de herdabilidade [$h^2 = q/(1+q)$] na segunda casa decimal (Graser et al., 1987).

9.6.2. Estimadores DF-REML para modelos com mais de um efeito aleatório

Modelo incluindo efeito aditivo (a) e um efeito aleatório adicional (c)

Para modelos que envolvem dois fatores aleatórios (além do resíduo) e três componentes de variância a serem estimados, Meyer (1989b) sugere a avaliação da razão θ entre o componente de variância e a variância fenotípica (soma-tória dos três componentes de variância) e não o parâmetro q (que é dado em relação a variância ambiental). Dessa forma, no presente caso necessitam ser avaliados os parâmetros θ_a e θ_c , relativos a σ_a^2 e σ_c^2 , respectivamente.

A função L a ser avaliada, no presente caso, equivale a:

$$L = -\frac{1}{2} \left[(N - r(x) - N_a - N_c) \log_e \hat{\sigma}_e^2 + \log_e |C| + N_a \log_e \hat{\sigma}_a^2 + \log_e |A| + N_c \log_e \sigma_c^2 + y' Py / \sigma_e^2 \right], \text{ em que:}$$

N_c : número de níveis no efeito aleatório c.

Neste caso, V (em P), refere-se a todos os efeitos aleatórios. Também $\log_e |A|$ não depende dos parâmetros a serem estimados, ou seja, é constante e não necessita ser calculada para maximizar L .

Modelo incluindo efeito aditivo (a), de dominância (d) e um efeito aleatório adicional (c)

Neste caso, necessitam ser avaliados os parâmetros θ_a , θ_d e θ_c , relativos a σ_a^2 , σ_d^2 e σ_c^2 , respectivamente. Considerando normalidade, para a determinação do máximo da função de verossimilhança ou equivalentemente, -2 vezes o seu logaritmo, basta avaliar a função:

$$\begin{aligned} -2 \log_e L = & \text{const} + y' Py / \sigma_e^2 + \log_e |C| + [(N - r(X) - 2N_a - N_c)] \log_e \sigma_e^2 + \\ & + N_a (\log_e \sigma_a^2 + \log_e \sigma_d^2) + N_c (\log_e \sigma_c^2) + \log_e |A| + \log_e |D| \end{aligned}$$

Uma vez que $\log_e |A|$ e $\log_e |D|$ não dependem dos parâmetros a serem estimados, não necessitam ser calculados com a finalidade de maximizar L .

9.7. Modelo Geral de Estimação REML e Predição BLUP

Um modelo linear misto geral é da forma (Henderson, 1984): $y = Xb + Za + e$ (1), com as seguintes distribuições e estruturas de médias e variâncias:

$$a \sim N(0, G)$$

$$E(y) = Xb$$

$$e \sim N(0, R)$$

$$\text{Var}(y) = V = ZGZ + R$$

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.

0 : vetor nulo.

Considerando 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:

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

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 ;

$\hat{a} = GZV^{-1}(y - X\hat{b}) = CV^{-1}(y - X\hat{b})$: 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 elas associadas 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:

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

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

em que:

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

$v = N - r(x)$: 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 Grazer et al. (1987); (iii) “Average Information” (AI) de Johnson & Tompson (1995). Estes algoritmos geraram as denominações EMREML, DFREML e AIREML.

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: vetor de efeitos genéticos aditivos.

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

σ_a^2 : variância genética aditiva.

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

σ_e^2 : variância residual.

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

$$y = Xb + Za + e$$

$$= Xb + Z_1 a^* + Z_2 p + e, \text{ em que:}$$

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

σ_p^2 : variância dos efeitos permanentes.

(c) Modelo multivariado, ajustando os efeitos aditivos

No caso bivariado, têm-se:

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

$$G = A \otimes G_0; \quad R = I \otimes R_0;$$

$$G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad R_0 = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma_{e_2}^2 \end{bmatrix} \quad \text{ou} \quad R_0 = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}, \text{ em que:}$$

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

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

(d) Modelo geoestatístico ou de séries temporais para análise espacial

$R = \Sigma$: matriz não diagonal que considera a correlação entre resíduos, por exemplo, linhas auto-regressivas e colunas auto-regressivas, ou covari-

ância baseada em semivariâncias ajustadas, para contemplar a autocorrelação espacial entre as observações.

Os modelos espaciais permitem estudar a variabilidade espacial do solo nas áreas experimentais, por meio 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. Maiores detalhes e exemplos de análise espacial de experimentos são apresentados no tópico 14.19.

9.8. Estimação REML de Componentes de Variância e Predição BLUP de Valores Genéticos Intrapopulacionais

Os valores genéticos aditivos intrapopulacionais preditos são úteis tanto para o melhoramento intrapopulacional (intraespecífico) quanto para o melhoramento interpopulacional (interespecífico). Para o melhoramento intrapopulacional, esses valores norteiam a seleção visando ao melhoramento progressivo da espécie. Para o melhoramento interpopulacional, permitem a escolha criteriosa dos indivíduos a serem empregados nos cruzamentos, bem como conduzem 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).

9.8.1. Delineamento em blocos ao acaso, progênies de polinização aberta, uma só população

9.8.1.1. 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:

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

$$G = A \sigma_a^2$$

$$R = I \sigma_e^2$$

$$C = I \sigma_c^2$$

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

Equações de modelo misto

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

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

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2}$: 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.

σ_a^2 : variância genética aditiva.

σ_c^2 : variância entre parcelas.

σ_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 equivalem às previsões obtidas pelo índice multiefeitos (Resende & Higa, 1994a), dado por:

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

$$b_1 = \frac{(1 - \rho_a) \sigma_a^2}{\sigma_{dp}^2}; \quad b_2 = \frac{\left[1 + (nb - 1) \rho_a\right] \sigma_a^2}{\sigma_{fa}^2 + \sigma_c^2 / b + \sigma_{dp}^2 / nb}; \quad b_3 = \frac{(1 - \rho_a) \sigma_a^2}{\sigma_{dp}^2 / n + \sigma_c^2}; \quad \rho_a: \text{ correlação genética aditiva intraclass} (\rho_a = 0,25 \text{ para famílias de meios-irmãos}).$$

σ_{fa}^2 , σ_c^2 e σ_{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

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

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

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

C^{22} e C^{33} advêm 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 o 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.

9.8.1.2. 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)

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

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, em termos computacionais, desejável e necessário. Os efeitos temporários da interação progênies x medições e progênies x medições x blocos são incorporados ao vetor 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

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

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

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

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

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

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

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

ou seja:

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

$$P = I \sigma_p^2$$

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

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 de 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 ao mesmo tempo como arquivo de dados e como arquivo de pedigree. Devem ser executados, de forma seqüencial, os subprogramas DFPREP e DFUNI.

9.8.1.3. 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 (ℓ) locais (experimentos) com algumas progênies 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

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

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

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

Para o caso que envolve três experimentos, tem-se:

$$\text{Var} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ c_1 \\ c_2 \\ c_3 \\ e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_1 a_2} & A\sigma_{a_1 a_3} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_1 a_2} & A\sigma_{a_2}^2 & A\sigma_{a_2 a_3} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_1 a_3} & A\sigma_{a_2 a_3} & A\sigma_{a_3}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{c_3}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_3}^2 \end{bmatrix}, \text{ em que:}$$

$\sigma_{a_1}^2$, $\sigma_{a_2}^2$ 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

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

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

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

$$R^{-1} = R_0^{-1} \otimes I; \quad G^{-1} = G_0^{-1} \otimes A^{-1}; \quad C^{-1} = C_0^{-1} \otimes I$$

$$R_0 = \begin{bmatrix} \sigma_{e_1}^2 & 0 & 0 \\ 0 & \sigma_{e_2}^2 & 0 \\ 0 & 0 & \sigma_{e_3}^2 \end{bmatrix}; \quad G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} & \sigma_{a_{13}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 & \sigma_{a_{23}} \\ \sigma_{a_{13}} & \sigma_{a_{23}} & \sigma_{a_3}^2 \end{bmatrix}; \quad C_0 = \begin{bmatrix} \sigma_{c_1}^2 & 0 & 0 \\ 0 & \sigma_{c_2}^2 & 0 \\ 0 & 0 & \sigma_{c_3}^2 \end{bmatrix}$$

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

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

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

$\sigma_{y_i}^2$ = variância fenotípica individual 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 (de preferência este). Neste caso, a função geral a ser maximizada é aquela descrita no item 9.7.

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 subprograma 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 alternativa é a seleção de indivíduos para plantio 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 $y^* = Xb + Za^* + Wc^* + e$, e os estimadores e preditores equivalem àqueles apresentados no item 9.8.1.1, sendo que os arquivos devem ser montados conforme uma estrutura univariada.

Outra situação refere-se à avaliação da eficiência da seleção indireta ou mesmo a 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 9.8.1.2, porém, substituindo-se Tp por Sf^{**} e σ_p^2 por σ_f^{**2} . Neste caso, σ_f^{**2} estima $(1/4)$ de σ_{ae}^2 , ao passo que $\sigma_{a^{**}}^2$ equivale a σ_a^2 , ou seja, a variância genética aditiva livre da interação genótipo x ambiente.

Considerando os três modelos alternativos apresentados neste tópico, as equivalências para o caso balanceado são:

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

- (ii) $\sigma_{ae}^2 = \frac{1}{2}(\sigma_{a_1}^2 - \sigma_{a_2}^2) + (1 - \rho_{a_{12}}) \sigma_{a_1} \sigma_{a_2} = 4\sigma_{f^{**}}^2;$
- (iii) $\sigma_a^2 + \sigma_{ae}^2 = (\sigma_{a_1}^2 + \sigma_{a_2}^2)/2;$
- (iv) $\sigma_{a^*}^2 = \frac{2\sigma_a^2 + \sigma_{a_1}^2 + \sigma_{a_2}^2}{4} = \sigma_a^2 + (1/2) \sigma_{ae}^2 = \sigma_{a^{**}}^2 + (1/2) \sigma_{ae}^2;$
- (v) $\sigma_{c^{**}}^2 = (\sigma_{c_1}^2 + \sigma_{c_2}^2)/2;$
- (vi) $\sigma_{e^{**}}^2 = (\sigma_{e_1}^2 + \sigma_{e_2}^2)/2;$
- (vii) $\sigma_{c^*}^2 = \sigma_{c^{**}}^2 + (1/2) \sigma_{f^{**}}^2;$
- (viii) $\sigma_{e^*}^2 = \sigma_{e^{**}}^2.$

Em resumo, opções de modelo de acordo com os respectivos objetivos são:

- (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.

9.8.1.4. 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 que:

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

Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \text{ Var } \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{p_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{p_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}, \text{ em que:}$$

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

Equações de modelo misto

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

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

$$R^{-1} = R_0^{-1} \otimes I; \quad G^{-1} = G_0^{-1} \otimes A^{-1}; \quad C^{-1} = C_0^{-1} \otimes I; \quad P^{-1} = P_0^{-1} \otimes I$$

$$R_0 = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_0 = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad P_0 = \begin{bmatrix} \sigma_{p_1}^2 & 0 \\ 0 & \sigma_{p_2}^2 \end{bmatrix}$$

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

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

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 9.7.

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.

9.8.2. Delineamento em blocos ao acaso, progênies de polinização controlada, uma só população

9.8.2.1. 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

Modelo linear misto (modelo aditivo-dominante univariado)

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

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

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

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

σ_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:

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

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

Estimadores iterativos dos componentes de variância por REML via algoritmo EM

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

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

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

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

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

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

Para ajuste dos efeitos de dominância, estes 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, o qual 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:

Indivíduo Pai Mãe Bloco Parcela Família de Irmãos Germanos Variável 1 . . .**Variável n**

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.

9.8.2.2. 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

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

Equações de modelo misto

$$\begin{bmatrix} XX & XZ & XW & XS & XT \\ ZX & ZZ + A^{-1}\lambda_1 & ZW & ZS & ZT \\ WX & WZ & WW + I\lambda_2 & WS & WT \\ SX & SZ & SW & SS + I\lambda_3 & ST \\ TX & TZ & TW & TS & TT + I\lambda_4 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{f} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \\ Wy \\ Sy \\ Ty \end{bmatrix}, \text{ em que:}$$

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

Estimadores iterativos de componentes de variância por REML via algoritmo EM

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

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

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

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

$\hat{\sigma}_p^2 = [\hat{p}'\hat{p} + \hat{\sigma}_e^2 \operatorname{tr} C^{55}] / s$, 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.

9.8.2.3. 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

Modelo linear misto (modelo aditivo-dominante multivariado)

No caso bivariado:

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

Distribuições e estruturas de médias e variâncias

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

$\sigma_{f_{12}}$: covariância entre os efeitos de dominância associado as famílias de irmãos germanos nos ambientes 1 e 2 (equivale a $\frac{1}{4} \sigma_d^2$, livre da interação).

Equações de modelo misto

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

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

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

$$R_O = \begin{bmatrix} \sigma_{e1}^2 & 0 \\ 0 & \sigma_{e2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a12} \\ \sigma_{a12} & \sigma_{a2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c1}^2 & 0 \\ 0 & \sigma_{c2}^2 \end{bmatrix}; \quad F_O = \begin{bmatrix} \sigma_{f1}^2 & \sigma_{f12} \\ \sigma_{f12} & \sigma_{f2}^2 \end{bmatrix}$$

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 9.7.

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**. Deverem ser executados os subprogramas DFPREP e DXMUX.

9.8.2.4. 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

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

Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}S & X'R^{-1}T \\ ZR^{-1}X & ZR^{-1}Z + G^{-1} & ZR^{-1}W & ZR^{-1}S & ZR^{-1}T \\ WR^{-1}X & WR^{-1}Z & WR^{-1}W + C^{-1} & WR^{-1}S & WR^{-1}T \\ SR^{-1}X & SR^{-1}Z & SR^{-1}W & SR^{-1}S + F^{-1} & SR^{-1}T \\ TR^{-1}X & TR^{-1}Z & TR^{-1}W & TR^{-1}S & TR^{-1}T + P^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{f} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ ZR^{-1}y \\ WR^{-1}y \\ SR^{-1}y \\ TR^{-1}y \end{bmatrix}, \text{ em que:}$$

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

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

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 9.7.

Software

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

9.8.3. Delineamento em blocos ao acaso, progênies de polinização aberta, várias populações

9.8.3.1. 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

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

Adicionalmente, existem duas opções para 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 9.8.1.1.

Na situação (i) (a), pode-se, com vantagem, 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 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 9.8.1.1.

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

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

Modelo linear misto (modelo aditivo univariado, multipopulações)

$y = Xb + Za + Wc_1 + Qr + Uc_2 + e$, 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 relativos 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

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

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

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

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

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

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

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

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

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

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

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

Equações de modelo misto

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

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

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

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

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

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{c_1}^2 + \sigma_r^2 + \sigma_{c_2}^2 + \sigma_e^2}$: 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 devida ao ambiente comum da parcela referente a progênie (sub-parcelas).

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

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

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

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

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

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

$s_1, t \text{ e } s_2$ = número de parcelas referentes a progênie, número de procedências e de parcelas de procedência, respectivamente.

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

Indivíduo Pai Mãe Bloco Parcela₁ Procedência Parcela₂ 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 pelo 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 estudar comparativamente as herdabilidades e coeficientes de variação genética das várias populações, deve-se ajustar um modelo do tipo apresentado em 9.8.1.1. para cada população.

9.8.3.2. Avaliação de progênieis 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 9.8.3.1, 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 9.8.1.2.
- 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, os estimadores e os preditores apresentados no tópico 9.8.1.2 devem ser empregados.

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

Modelo linear misto (modelo aditivo univariado, multipopulações, de reproductibilidade)

$y = Xb + Za + Wc_1 + Tp + Qr + Uc_2 + e$, 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

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

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

Assim:

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

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

Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q & X'U \\ ZX & ZZ + A^{-1}\lambda_1 & ZW & ZT & ZQ & ZU \\ WX & WZ & WW + I\lambda_2 & WT & WQ & WU \\ TX & TZ & TW & TT + I\lambda_3 & TQ & TU \\ QX & QZ & QW & QT & Q'Q + I\lambda_4 & QU \\ UX & UZ & UW & UT & UQ & UU + I\lambda_5 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{p} \\ \hat{r} \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \\ Wy \\ Ty \\ Q'y \\ Uy \end{bmatrix}, \text{ em que:}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Indivíduo Pai Mãe Bloco Parcela₁ Permanente Procedência Parcela₂ Variável 1 . . . Variável n

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

Na situação (ii) (b), a interação aleatória procedência x bloco tende a ser não significativa e o modelo, os preditores e os 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 (Ver tópicos 9.11 e 13.6).

Tanto na situação (i) (b) quanto na de (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 destacar que o próprio modelo de repetibilidade considera que a interação com medições seja, aproximadamente, nula (isto é, a correlação genética, 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.

9.8.3.3. 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

Para as situações descritas em (i) (a) e (ii) (a) do tópico 9.8.3.1, basta adotar o modelo, os estimadores e os preditores descritos no tópico 9.8.1.3. Para as situações (i)(b) ou (ii) (b), incluindo a interação procedência x bloco (efeito c_2), têm-se os modelos, estimadores e preditores descritos a seguir.

Modelo linear misto (modelo aditivo bivariado, multipopulações)

$y = Xb + Za + Wc_1 + Qr + Uc_2 + e$, 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 progênies, dos efeitos aleatórios de populações, dos efeitos aleatórios de parcelas relativos 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

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \text{Var} \begin{bmatrix} a_1 \\ a_2 \\ c_{11} \\ c_{12} \\ r_1 \\ r_2 \\ c_{21} \\ c_{22} \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_12} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_{11}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_{12}}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{r_1}^2 & I\sigma_{r_12} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{r_12} & I\sigma_{r_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{c_{21}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{c_{22}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

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

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

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

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

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

$$G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad \mathfrak{R}_O = \begin{bmatrix} \sigma_{r_1}^2 & \sigma_{r_{12}} \\ \sigma_{r_{12}} & \sigma_{r_2}^2 \end{bmatrix}; \quad C_{10} = \begin{bmatrix} \sigma_{c_{11}}^2 & 0 \\ 0 & \sigma_{c_{12}}^2 \end{bmatrix}; \quad C_{20} = \begin{bmatrix} \sigma_{c_{21}}^2 & 0 \\ 0 & \sigma_{c_{22}}^2 \end{bmatrix}; \quad R_O = \begin{bmatrix} \sigma_{\theta_1}^2 & 0 \\ 0 & \sigma_{\theta_2}^2 \end{bmatrix}$$

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:

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

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

$\rho_{a_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i} \sigma_{a_j}}$: 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 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 9.7.

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

Experimento Indivíduo Pai Mãe Bloco Parcela₁ Procedência Parcela₂ Variável 1 . . . Variável n

9.8.3.4. 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, avaliado em vários experimentos com algumas progêneres ou tratamentos comuns

Para as situações descritas em (i) (a) e (ii) (a) do tópico 9.8.3.1, basta adotar o modelo, estimadores e preditores descritos no tópico 9.8.1.4. Para as situações (i)(b) ou (ii) (b), excluindo a interação procedência x bloco (efeito c_2), tem-se os modelos, estimadores e preditores descritos a seguir.

Modelo linear misto (modelo aditivo bivariado, multipopulações, 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

$$\text{Var}(y_1) = V_1 = Z_1 A \sigma_{a_1}^2 Z_1 + W_1 I \sigma_{c_{11}}^2 W_1 + Q_1 I \sigma_{r_1}^2 Q_1 + T_1 I \sigma_{p_1}^2 T_1 + I \sigma_{e_1}^2$$

$$\text{Var}(y_2) = V_2 = Z_2 A \sigma_{a_2}^2 Z_2 + W_2 I \sigma_{c_{12}}^2 W_2 + Q_2 I \sigma_{r_2}^2 Q_2 + T_2 I \sigma_{p_2}^2 T_2 + I \sigma_{e_2}^2$$

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.

$\sigma_{c_{11}}^2$ 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.

$\sigma_{p_1}^2$: covariância genética entre os locais 1 e 2, ao nível do efeito de procedência.

$\sigma_{p_2}^2$: 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

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

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

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

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:

$\rho_{a_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i} \sigma_{a_j}}$: 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:

$$\sigma_{ae}^2 = \frac{1}{2}(\sigma_{ai} - \sigma_{aj})^2 + (1 - \rho_{aij}) \sigma_{ai} \sigma_{aj}.$$

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

$$\sigma_{re}^2 = \frac{1}{2}(\sigma_{ri} - \sigma_{rj})^2 + (1 - \rho_{rij}) \sigma_{ri} \sigma_{rj}.$$

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 9.7.

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

Experimento Indivíduo Pai Mãe Bloco Parcela Procedência Permanente Variável 1 . . . Variável n

9.8.4. Delineamento em látice, progênies de polinização aberta, uma população

9.8.4.1. 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

Modelo linear misto (modelo aditivo univariado, em látice)

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

y, b, a, c, ∂ e e: vetores de dados, dos efeitos de repetições (fixos), de efeitos genéticos aditivos (aleatórios), dos efeitos de 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 ∂ , 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 consideradas nulas.

Assim:

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

$$G = A \sigma_A^2$$

$$C = I \sigma_c^2$$

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

$$R = I \sigma_e^2$$

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

Equações de modelo misto

$$\begin{bmatrix} XX & X'Z & X'W & X'H \\ ZX & ZZ + A^{-1}\lambda_1 & ZW & ZH \\ WX & WZ & WW + I\lambda_2 & WH \\ HX & H'Z & H'W & HH + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{\partial} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Wy \\ Hy \end{bmatrix}, \text{ em que:}$$

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

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2 + \sigma_{\partial}^2}$: 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.

σ_a^2 : variância genética aditiva.

σ_c^2 : variância entre parcelas.

σ_{∂}^2 : variância entre blocos dentro de repetições.

σ_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

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

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

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

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

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

tr : operador traço matricial.

r(x) : posto da matriz X.

N, q, s, η : número total de dados, de indivíduos, de parcelas e 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 que funciona ao mesmo tempo como arquivo de pedigree e como arquivo de dados. Deve ser executado o subprograma DFPREP e, em seguida, o DFUNI.

9.8.4.2. 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

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

$y = Xb + Za + Wc + Tp + H\partial + e$, 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, desejável e necessário do ponto de vista computacional.

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 consideradas nulas.

Assim:

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

$$P = I \sigma_p^2$$

$$V = ZA\sigma_a^2 Z' + Wl\sigma_c^2 W' + Tl\sigma_p^2 T' + Hl\sigma_\partial^2 H' + l\sigma_e^2.$$

Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'T & X'H \\ ZX & ZZ + A^{-1}\lambda_1 & ZW & ZT & ZH \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'H \\ T'X & TZ & TW & T'T + I\lambda_3 & TH \\ H'X & H'Z & HW & HT & H'H + I\lambda_4 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \\ \hat{\partial} \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \\ Wy \\ Ty \\ Hy \end{bmatrix}, \text{ em que:}$$

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

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_p^2 + \sigma_e^2 + \sigma_\partial^2}$: 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

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

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

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

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

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

C^{22}, C^{33}, C^{44} e C^{55} advêm 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, de modo seqüencial, os subprogramas DFPREP e DFUNI.

9.8.4.3. 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 (ℓ) experimentos com algumas progêneres ou tratamentos comuns

Modelo linear misto (modelo aditivo multivariado, em látice)

No caso bivariado, têm-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

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \text{ Var } \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_2} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_2} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{\hat{\sigma}_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{\hat{\sigma}_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}, \text{ em que:}$$

$\sigma_{\hat{\sigma}_1}^2$ e $\sigma_{\hat{\sigma}_2}^2$. = variância permanente entre blocos dentro de repetições nos locais 1 e 2, respectivamente.

Equações de modelo misto

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

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

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

$$R_O = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_O = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 a_2} \\ \sigma_{a_1 a_2} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_O = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad B_O = \begin{bmatrix} \sigma_{\hat{\sigma}_1}^2 & 0 \\ 0 & \sigma_{\hat{\sigma}_2}^2 \end{bmatrix}$$

Os componentes de variância estão associados aos parâmetros h^2 , c^2 , $\hat{\sigma}^2$ e ρ_{a_i} , da seguinte maneira:

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{e_i}^2 = (1 - \rho_i) \sigma_{y_i}^2; \quad \sigma_{c_i}^2 = c_i^2 \sigma_{y_i}^2; \quad \sigma_{\hat{\sigma}_i}^2 = (\hat{\sigma}_i^2 - h_i^2 - c_i^2) \sigma_{y_i}^2; \quad \sigma_{a_j} = \rho_{a_j} \sigma_{a_i} \sigma_{a_j}$$

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 9.7.

Seqüência de colunas de dados para análise no software DFREML

Experimento Indivíduo Pai Mãe Repetição Parcela Bloco

Variável 1 . . . Variável n

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

9.8.4.4. 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

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

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

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Estruturas de médias e variâncias

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ \delta_1 \\ \delta_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \text{Var} \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ \delta_1 \\ \delta_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{p_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{p_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{\delta_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{\delta_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

Equações de modelo misto

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}T & X'R^{-1}H \\ ZR^{-1}X & ZR^{-1}Z+G^{-1} & ZR^{-1}W & ZR^{-1}T & ZR^{-1}H \\ WR^{-1}X & WR^{-1}Z & WR^{-1}W+C^{-1} & WR^{-1}T & WR^{-1}H \\ TR^{-1}X & TR^{-1}Z & TR^{-1}W & TR^{-1}T+P^{-1} & TR^{-1}H \\ HR^{-1}X & HR^{-1}Z & HR^{-1}W & HR^{-1}T & HR^{-1}H+P^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ ZR^{-1}y \\ WR^{-1}y \\ TR^{-1}y \\ HR^{-1}y \end{bmatrix}, \text{ em que:}$$

$$R^{-1} = R_0^{-1} \otimes I; \quad G^{-1} = G_0^{-1} \otimes A^{-1}; \quad C^{-1} = C_0^{-1} \otimes I; \quad P^{-1} = P_0^{-1} \otimes I; \quad B^{-1} = B_0^{-1} \otimes I$$

$$R_0 = \begin{bmatrix} \sigma_{\theta_1}^2 & 0 \\ 0 & \sigma_{\theta_2}^2 \end{bmatrix}; \quad G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_0 = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad P_0 = \begin{bmatrix} \sigma_{p_1}^2 & 0 \\ 0 & \sigma_{p_2}^2 \end{bmatrix}; \quad B_0 = \begin{bmatrix} \sigma_{\delta_1}^2 & 0 \\ 0 & \sigma_{\delta_2}^2 \end{bmatrix}$$

Estimadores dos componentes de variância por REML

Neste caso, é recomendado o algoritmo AI. A função geral a ser maximizada é descrita no item 9.7.

Software

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

9.8.5. Delineamento em látice, progênies de polinização aberta, várias populações

9.8.5.1. 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

Para as situações descritas em (i) (a) e (ii) (a) do tópico 9.8.3.1, basta adotar o modelo, os estimadores e os preditores descritos no tópico 9.8.4.1. Para as situações (i)(b) ou (ii) (b), incluindo a interação procedência x bloco (efeito c_2), os modelos, estimadores e preditores são descritos a seguir.

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

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

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

∂ = vetor de efeitos aleatórios dos blocos dentro de repetições.

H = matriz de incidência para ∂ .

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 consideradas nulas.

Assim:

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

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

Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'H & X'Q & X'U \\ Z'X & ZZ + A^{-1}\lambda_1 & Z'W & Z'H & Z'Q & Z'U \\ W'X & W'Z & WW + I\lambda_2 & W'H & W'Q & W'U \\ H'X & H'Z & H'W & HH + I\lambda_3 & H'Q & H'U \\ Q'X & Q'Z & Q'W & Q'H & Q'Q + I\lambda_4 & Q'U \\ U'X & U'Z & U'W & U'H & U'Q & U'U + I\lambda_5 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{\partial} \\ \hat{r} \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \\ Wy \\ Hy \\ Qy \\ Uy \end{bmatrix}, \text{ em que:}$$

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

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

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

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

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

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

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

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

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

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

Estimadores iterativos dos componentes de variância por REML via algoritmo EM

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

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

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

$$\hat{\sigma}_{\hat{\theta}}^2 = [\hat{\theta}' \hat{\theta} + \hat{\sigma}_e^2 \operatorname{tr}(C^{44})] / \eta$$

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

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

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

Indivíduo Pai Mãe Repetição Parcela₁ Bloco Procedência Parcela₂ 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.

9.8.5.2. 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

Considerando-se as situações (i) (a), (i) (b), (ii) (a) e (ii) (b) descritas no tópico 9.8.3.1, 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 9.8.4.2.
- 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 9.8.4.2. devem ser empregados.

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

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

$y = Xb + Za + Wc_1 + Tp + Qr + H\hat{\partial} + e$, 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

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

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

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

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

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

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

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

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

Assim:

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

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

Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q & X'H \\ ZX & ZZ + A^{-1}\lambda_1 & ZW & ZT & ZQ & ZH \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'Q & W'H \\ TX & TZ & TW & TT + I\lambda_3 & TQ & TH \\ Q'X & Q'Z & Q'W & QT & Q'Q + I\lambda_4 & Q'H \\ H'X & H'Z & HW & HT & HQ & HH + I\lambda_5 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c}_1 \\ \hat{p} \\ \hat{r} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \\ Wy \\ Ty \\ Qy \\ Hy \end{bmatrix}, \text{ em que:}$$

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

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

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

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

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

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

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

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

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

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

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

Estimadores iterativos dos componentes de variância por REML via algoritmo EM

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

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

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

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

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

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

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

Indivíduo Pai Mãe Repetição Parcela₁ Permanente Procedência Bloco Variável 1 . . . Variável n

É preciso apenas um arquivo, que funciona ao mesmo tempo como de pedigree e de dados. Primeiro, deve ser executado o subprograma DFPREP e, depois, 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.

9.8.6. Avaliação apenas de populações

9.8.6.1. 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, multipopulações, 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:

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

$$G = A \sigma_A^2$$

$$R = I \sigma_c^2$$

$$C = I \sigma_e^2$$

$$R = I \sigma_r^2$$

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

Equações de modelo misto

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

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

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_r^2 + \sigma_e^2}$: 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

Indíduo Pai Mãe Bloco Parcela Procedência Variável 1 . . . Variável n

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

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

9.8.7. Avaliação simultânea de caracteres

9.8.7.1. Avaliação de progêneres 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

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

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

$$R^{-1} = R_0^{-1} \otimes I; \quad G^{-1} = G_0^{-1} \otimes A^{-1}; \quad C^{-1} = C_0^{-1} \otimes I;$$

$$R_0 = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma_{e_2}^2 \end{bmatrix}; \quad G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{11}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_0 = \begin{bmatrix} \sigma_{c_1}^2 & \sigma_{c_{12}} \\ \sigma_{c_{12}} & \sigma_{c_2}^2 \end{bmatrix};$$

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

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{c_i}^2 = c_i^2 \sigma_{y_i}^2; \quad \sigma_{\theta_i}^2 = (1 - h_i^2 - c_i^2) \sigma_{y_i}^2; \quad \sigma_{a_j} = \rho_{a_j} \sigma_{a_i} \sigma_{a_j}, \text{ em que:}$$

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

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 9.7.

Seqüência de colunas de dados para análise no software DFREML

Caráter 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 estimação e predição simultânea de caracteres utilizam os mesmos modelos apresentados nos itens 9.8.1.3, 9.8.1.4, 9.8.2.3, 9.8.2.4, 9.8.3.3, 9.8.3.4, 9.8.4.3 e 9.8.4.4, 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 concomitante 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.

9.8.8. Espécies com sistema reprodutivo misto

Os modelos, estimadores e preditores apresentados nos itens 9.8.1.1, 9.8.1.2, 9.8.1.3, 9.8.1.4, 9.8.3.1, 9.8.3.2, 9.8.3.3, 9.8.3.4, 9.8.4.1, 9.8.4.2, 9.8.4.3, 9.8.4.4, 9.8.5.1 e 9.8.5.2 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. (1995c). Posteriormente, as herdabilidades (ou variâncias e covariâncias genéticas) corrigidas devem ser fixadas nos programas computacionais, para a predição dos valores genéticos. Uma abordagem mais completa sobre avaliação genética sob sistema reprodutivo misto é apresentada no capítulo 18.

9.9. Estimação REML de Componentes de Variância e Predição BLUP de Valores Genéticos Interpopulacionais

9.9.1. Híbridos incluindo duas espécies ou populações

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

$$\begin{aligned} G_{(12)} &= \mu_{(12)} + a_{(12)} + d_{(12)} \\ &= \frac{1}{2}(\mu_1 + \mu_2) + h_{(12)} + \frac{1}{2}(a_{12} + a_{21}) + \delta_{(12)} + m_{(12)}, \text{ em que:} \end{aligned}$$

$\mu_{(12)}$ = média geral da população híbrida.

$a_{(12)}$ = efeito genético aditivo de um indivíduo híbrido.

$d_{(12)}$ = efeito de dominância de um indivíduo híbrido.

μ_1 e μ_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 μ_1 , μ_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 estas e dos efeitos genéticos aditivos interpopulacionais dos genitores a serem cruzados, pode-se predizer a descendência híbrida. Se a correlação entre os efeitos gené-

ticos 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 de (\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 progênieis 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

É preciso avaliar as progênieis de irmãos germanos interpopulacionais obtidas sob cruzamentos fatoriais, dialélicos ou hierárquicos ou a realização de testes clonais com indivíduos híbridos. Neste caso, 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 progênieis são utilizados:

(a) Apenas avaliação de progênieis interpopulacionais

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

Tem-se o seguinte modelo linear misto (modelo univariado aditivo-dominante).

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

$y, b, a_{(12)}, d_{(12)}, c, e$ = 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

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

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

$\sigma_{a_{(12)}}^2$ = variância genética aditiva interpopulacional.

$\sigma_{d_{(12)}}^2$ = variância de dominância interpopulacional.

σ_c^2 = variância entre parcelas.

σ_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

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

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

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

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

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

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

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

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

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

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:

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

As matrizes de parentesco A e D são da forma:

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

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 equivalem a:

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

Considerando $\sigma_{a_{(12)}}^2 = 2,75$, $\sigma_{d_{(12)}}^2 = 1,00$ e $\sigma_e^2 = 3,125$, têm-se $h_{a_{(12)}}^2 = 0,4$, $h_{g_{(12)}}^2 = 0,55$, $\lambda_1 = 1,1364$ e $\lambda_2 = 3,125$. Resolvendo-se as equações de modelo misto obtém-se:

Efeitos	Soluções
Blocos	
\hat{b}_1	19,8667
\hat{b}_2	20,1803
Efeitos genéticos aditivos	
\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
Efeitos de dominância	
\hat{d}_1	0
\hat{d}_2	0

Continua...

Efeitos	Soluções
\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, 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

Os valores genéticos aditivos 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 por meio de plantios por sementes

O valor genotípico predito da descendência do cruzamento entre os genitores 1 e 2 é dado por $\hat{G}_{(12)} = \hat{\mu}_{(12)} + \frac{1}{2}(\hat{a}_1 + \hat{a}_2) + \hat{\delta}_{(12)}$. No caso, $\delta_{(12)}$ 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:

$$\begin{aligned}\hat{G}_{(12)} &= \hat{\mu}_{(12)} + \frac{1}{2}(\hat{a}_1 + \hat{a}_2) + \frac{1}{2}(\hat{d}_7 + \hat{d}_8) \\ &= 20,0235 + \frac{1}{2}(1,2865 + 1,2865) + \frac{1}{2}(0,7929 + 0,3767) \\ &= 21,8948 \text{ metros}\end{aligned}$$

(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 $\delta_{(12)}$ a partir de cruzamentos dialélicos, fatoriais ou hierárquicos

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

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

$\mathbf{y} = \mathbf{Xb} + \mathbf{Za}_{(12)} + \mathbf{Wc} + \mathbf{S}\delta_{(12)} + \mathbf{e}$, em que:

$\delta_{(12)}$ é o efeito aleatório de dominância associado à família de irmãos germanos.

\mathbf{S} é a matriz de incidência para $\delta_{(12)}$.

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

Estrutura de médias de variâncias

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

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

Equações de modelo misto

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

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

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 progênies interpopulacionais e intrapopulacionais

Muitas vezes são geradas, simultaneamente, progênies intrapopulacionais e interpopulacionais de cada matriz. Este esquema é vantajoso, pois permite, adicionalmente, a obtenção de informações sobre a correlação genética (ρ_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 que trate os materiais genéticos intrapopulacionais e interpopulacionais como caracteres distintos. Isto é justificável quando ρ_a não equivale 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)

$$y = Xb + Za + Wc + S\delta + e$$

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

y_{11}, y_{12}, y_{22} : vetores de dados correspondentes às progêñies 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.

$\delta_{11}, \delta_{12}, \delta_{22}$: 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

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

$\sigma_{a_{11}}^2$, $\sigma_{a_{12}}^2$ 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.

$\sigma_{\delta_{122}}^2$: 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 $\sigma_{a_{1122}} = 0$, pode-se optar, também, pelo ajuste de dois modelos bivariados (um envolvendo y_{11} e y_{12} e outro, y_{12} e y_{22}) em vez do modelo trivariado.

A partir do modelo apresentado, podem ser obtidas 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

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

$$R^{-1} = R_0^{-1} \otimes I; \quad G^{-1} = G_0^{-1} \otimes A^{-1}; \quad C^{-1} = C_0^{-1} \otimes I; \quad \Delta^{-1} = \Delta_0^{-1} \otimes I$$

$$R_0 = \begin{bmatrix} \sigma_{e_{11}}^2 & 0 & 0 \\ 0 & \sigma_{e_{12}}^2 & 0 \\ 0 & 0 & \sigma_{e_{22}}^2 \end{bmatrix}; \quad G_0 = \begin{bmatrix} \sigma_{a_{11}}^2 & \sigma_{a_{112}}^2 & 0 \\ \sigma_{a_{112}}^2 & \sigma_{a_{12}}^2 & \sigma_{a_{22}}^2 \\ 0 & \sigma_{a_{122}}^2 & \sigma_{a_{22}}^2 \end{bmatrix}; \quad C_0 = \begin{bmatrix} \sigma_{c_{11}}^2 & 0 & 0 \\ 0 & \sigma_{c_{12}}^2 & 0 \\ 0 & 0 & \sigma_{c_{22}}^2 \end{bmatrix}; \quad \Delta_0 = \begin{bmatrix} \sigma_{\delta_{11}}^2 & 0 & 0 \\ 0 & \sigma_{\delta_{12}}^2 & 0 \\ 0 & 0 & \sigma_{\delta_{22}}^2 \end{bmatrix}$$

Estimadores dos componentes de variância por REML

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 9.7.

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

População Indivíduo Pai Mãe Bloco Parcела 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.

9.9.2. Híbridos envolvendo três ou mais espécies

Híbridos de três ou mais espécies têm sido comuns no melhoramento do *Eucalyptus* no Brasil (Assis, 2000). Nesta situação, são considerados na análise vários tipos de híbridos. Por exemplo, um híbrido triplo de *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 implicado em vários cruzamentos interpopulacionais de forma que uma avaliação completa deve englobar 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 contendo as várias combinações de híbridos entre duas espécies e as combinações das três espécies.

Modelo linear misto (modelo univariado de heterose fixa)

$y = Xb + Th + Za + Wc + S\delta + e$, 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.

δ : 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 δ , respectivamente.

Neste caso, os valores genotípicos preditos de cada cruzamento visando explorar a capacidade específica de combinação é dado por $\hat{g} = Xb + Th + Za + S\delta$, em que \bar{a} refere-se à média dos efeitos aditivos interpulacionais dos genitores considerados 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 9.9.1 (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.

9.10. Componentes de Variância e Valores Genotípicos em Testes Clonais

9.10.1. 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 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 e W : matrizes de incidência para b, g e c , respectivamente.

Distribuições e estruturas de médias e variâncias

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

Equações de modelo misto

$$\begin{bmatrix} X'X & X'Z & X'W \\ ZX & ZZ + I_{\lambda_1} & ZW \\ W'X & W'Z & WW + I_{\lambda_2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \\ Wy \end{bmatrix}, \text{ em que:}$$

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

$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_c^2 + \sigma_e^2}$: herdabilidade individual no sentido amplo no bloco.

σ_g^2 : variância genotípica entre clones.

σ_c^2 : variância entre parcelas.

σ_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 parentados. 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.

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

Modelo linear misto (modelo univariado aditivo-dominante)

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

d: vetor aleatório dos efeitos de dominância.

Distribuições e estruturas de médias e variâncias

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

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

Equações de modelo misto

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

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

σ_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:

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

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

Estimadores iterativos dos componentes de variância por REML via algoritmo EM

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

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

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

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

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, estes 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. Esse 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 relativo 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.

9.11. Escolha e Comparação entre Modelos de Análise

Além de sua utilidade na estimação, o princípio da verossimilhança também permite comparar a adequabilidade de vários modelos, desde que tenham uma estrutura hierárquica ou aninhada. Dados dois modelos U e V com máximos das funções de verossimilhança (restrita ou não) $L(U)$ e $L(V)$ e correspondentes números de parâmetros n_u e n_v , tem-se que menos duas vezes o logaritmo da razão de verossimilhança, $D = -2 \log_e (L(U)/L(V))$, possui, aproximadamente, distribuição χ^2 com $n_v - n_u$ graus de liberdade (assumindo U como hierárquico ou um caso especial de V).

As funções L e $\log_e L$ apresentam o mesmo máximo. Estas funções podem ser expressas também em termos da função $-2 \log_e L$, que deve ser minimizada. Para maximização de L ou $\log_e L$, valores iniciais dos parâmetros são fornecidos e a cada iteração, novos valores são calculados e substituídos na função $-2 \log_e L$ até que esta seja minimizada (e, de maneira equivalente, $\log_e L$ será maximizada) e os parâmetros converjam para os valores constantes.

A significância da diferença no ajuste de diferentes modelos aos dados pode ser testada usando o Teste da Razão de Verossimilhança – LRT, definido por:

$$\lambda = 2[\log_e L_{p+1} - \log_e L_p].$$

Assim, basta comparar λ (2 vezes a diferença (modelo com maior número de parâmetros – modelo com menor número de parâmetros) de $\log_e L$ associa-

dos a dois modelos ajustados) com o valor da função densidade de probabilidade (Tabela de χ^2) para determinado número de graus de liberdade e probabilidade de erro (Dobson, 1990). O número de graus de liberdade é definido pela diferença no número de parâmetros ou componentes de variância entre modelos.

9.12. Estimação da Precisão das Estimativas dos Componentes de Variância

Os erros das estimativas dos componentes de variância podem ser obtidos, aproximadamente, pela inversa da matriz das derivadas segundas (matriz Hessiano) da função $\log_e L$ com respeito aos componentes de variância (parâmetros a serem estimados). A matriz Hessiano não é obtida com o algoritmo DFREML, mas aproximações por diferenciação numérica podem ser obtidas (Nelder & Mead, 1965), as quais são adequadas para modelos com até dois efeitos aleatórios (Meyer, 1988). Com o advento do algoritmo AI, os erros (aproximados) dos componentes de variância podem ser obtidos da matriz de informação média (Meyer, 1998).

O teste da razão de verossimilhança – LRT – pode também ser utilizado para testar duas hipóteses de interesse: (i) se determinada estimativa (da herdabilidade, por exemplo) difere de zero; (ii) se determinada estimativa difere de uma estimativa média para o caráter, obtida de literatura. Para a realização destes testes é necessário fixar nos programas computacionais (DFREML e MTDFREML, por exemplo) de estimação de componentes de variância/predição de valores genéticos, os valores de herdabilidade contra os quais se quer testar (no caso, zero ou o valor médio de literatura). A seguir aplica-se o LRT, comparando os modelos com a herdabilidade fixada a priori e com a herdabilidade estimada livremente com os próprios dados experimentais. Por exemplo, usando os resultados de Roman et al. (2000), um $\log_e L(1) = -2669,9062$ associado à estimativa da herdabilidade (estimada em 0,26) a partir dos próprios dados experimentais pode ser comparado com um $\log_e L(2) = -2674,6676$ obtido com a herdabilidade fixada em zero e também com um $\log_e L(3) = -2969,9088$ obtido com a herdabilidade fixada em 0,25. Neste caso, $2[\log_e L(1) - \log_e L(2)] = 9,52$, valor este significativo pelo teste qui-quadrado e que revela que a estimativa obtida (0,26) é diferente de zero. Por outro lado, $2[\log_e L(1) - \log_e L(3)] = 0,01$, valor este não significativo, revelando que a estimativa obtida (0,26) não difere significativamente do valor médio (0,25) de literatura para o caráter.

9.13. Exemplo Prático pelo DFREML

Como exemplo, considere a avaliação de 41 indivíduos (Eucalipto) para a variável altura aos 8 meses de idade, conforme a planilha apresentada a seguir.

Indivíduo	Pai	Mãe	População	Bloco	Altura (m)
6	1	0	1	1	2,24
7	1	0	1	1	2,65
8	1	0	1	2	2,41
9	1	0	1	2	2,25
10	1	0	1	3	2,58
11	1	0	1	3	2,67
12	1	0	1	3	2,71
13	1	0	1	3	2,47
14	2	0	1	2	2,29
15	2	0	1	2	2,26
16	2	0	1	3	1,97
17	2	0	1	3	2,14
18	2	0	1	3	2,44
19	2	0	1	3	2,52
20	2	0	1	3	1,72
21	2	0	1	3	2,75
22	3	0	1	1	2,38
23	3	0	1	2	2,46
24	3	0	1	3	2,29
25	3	0	1	3	2,30
26	3	0	1	3	2,94
27	4	0	2	1	2,50
28	4	0	2	1	2,44
29	4	0	2	2	2,44
30	4	0	2	2	2,15
31	4	0	2	3	2,54
32	4	0	2	3	2,74
33	4	0	2	3	2,50
34	4	0	2	3	2,54
35	5	0	2	1	2,65
36	5	0	2	2	2,52
37	5	0	2	2	2,67
38	5	0	2	3	2,79
39	5	0	2	3	2,33
40	5	0	2	3	2,67
41	5	0	2	3	2,69

Da forma como está representada, essa planilha pode ser utilizada diretamente na análise pelo software DFREML (Meyer, 1998), cujo arquivo correspondente contém a seguinte seqüência: três primeiras colunas referentes ao pedigree (com a numeração de identificação dos pais e mães devendo ser obrigatoriamente menor que a numeração dos indivíduos); as duas colunas seguintes são referentes aos efeitos fixos do modelo; a última coluna é relativa à variável a ser analisada.

Atribuindo-se o valor 0,5 como estimativa inicial para a herdabilidade e permitindo a busca variar no intervalo 0,30 – 0,70, são encontradas as seguintes estimativas em cada avaliação da função até a convergência:

Nº da avaliação	\hat{h}^2	$\hat{\sigma}_e^2$	Log L
2	0,7000	0,0183	25,8832
3	0,9000	0,0064	25,8380
4	0,6404	0,0216	25,8870
5	0,6335	0,0221	25,8871
6	0,6303	0,0223	25,8871
7	0,6298	0,0223	25,8871
8	0,6296	0,0223	25,8871
9	0,62959	0,0223	25,8871
10	0,62958	0,0223	25,8871
11	0,62958	0,0223	25,8871

Assim, após 11 avaliações da função, obteve-se a convergência para o valor de herdabilidade.

9.14. Estimação de Efeitos e Componentes de Variância Epistáticos

9.14.1. Efeitos genéticos e suas variâncias

O valor fenotípico de um caráter quantitativo em uma população panmítica é dado por $y = \mu + a + d + aa + ad + dd + aaa + e$ (1), ignorando as interações interlocos (epistasia) de maior ordem, em que:

μ :média genotípica.

a :efeito genético aditivo.

d :efeito de dominância.

aa :efeito da interação epistática do tipo aditivo x aditivo, entre dois locos.

ad :efeito da interação epistática do tipo aditivo x dominante, entre dois locos.

dd : efeito da interação epistática do tipo dominante x dominante, entre dois locos.

aaa : efeito da interação epistática do tipo aditivo x aditivo x aditivo, entre três locos.

e :efeito ambiental.

Com base no modelo para os valores fenotípicos, a variabilidade fenotípica pode ser decomposta em:

$\sigma_y^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{aa}^2 + \sigma_{ad}^2 + \sigma_{dd}^2 + \sigma_{aaa}^2 + \sigma_e^2$ (2), em que σ_i^2 refere-se à variância do efeito i ($i = y, a, d, aa, ad, dd, aaa$ e e).

A estimativa dos componentes da variação fenotípica e parâmetros genéticos associados baseia-se na utilização de dados fenotípicos associados a vários grupos de indivíduos com alguma relação de parentesco. Tais grupos de indivíduos geralmente constituem uma família e as relações de parentesco entre

os indivíduos dentro de famílias são utilizadas na estimação dos componentes de variância genotípica.

9.14.2. Covariância entre parentes e estimação de variâncias epistáticas

Em espécies alógamas, as principais relações de parentesco presentes nos dados de campo associados aos programas de melhoramento são apresentadas a seguir, em termos de covariância genética entre parentes.

A expressão geral para a covariância (cov) entre os parentes x e y é dada por (Cockerham, 1954):

$$\text{cov}_{(x,y)} = \alpha\sigma_a^2 + \delta\sigma_d^2 + \alpha^2\sigma_{aa}^2 + \alpha\delta\sigma_{ad}^2 + \delta^2\sigma_{dd}^2 + \alpha^3\sigma_{aaa}^2 + \dots \quad (3)$$

α : correlação genética aditiva entre os indivíduos x e y .

δ : correlação genética de dominância entre os indivíduos x e y .

A correlação genética de dominância entre um indivíduo x com genitores s e d e um indivíduo y com genitores m e f , em uma população não endógama é dada por $\delta = 0,25 (\alpha_{sm}\alpha_{df} + \alpha_{sd}\alpha_{mf})$, em que α_{ij} é a correlação genética aditiva entre os indivíduos i e j . Para irmãos germanos: $\delta = 0,25 [(1)(1) + (0)(0)] = 0,25$. Assim, todos os coeficientes em $\text{cov}_{(x,y)}$ podem ser obtidos a partir do parentesco aditivo entre x e y .

Com base na expressão (3), obtêm-se as seguintes correlações e covariâncias entre parentes:

(a) Parentesco de meios-irmãos:

$$\alpha = 1/4; \quad \delta = 0; \quad \alpha^2 = 1/16; \quad \alpha\delta = 0; \quad \delta^2 = 0; \quad \alpha^3 = 1/64 \quad (4)$$

$$\text{cov}_{(x,y)} = \alpha\sigma_a^2 + \alpha^2\sigma_{aa}^2 + \alpha^3\sigma_{aaa}^2 = (1/4) \sigma_a^2 + (1/16) \sigma_{aa}^2 + (1/64) \sigma_{aaa}^2$$

(b) Parentesco de irmãos germanos:

$$\alpha = 1/2; \quad \delta = 1/4; \quad \alpha^2 = 1/4; \quad \alpha\delta = 1/8; \quad \delta^2 = 1/16; \quad \alpha^3 = 1/8 \quad (5)$$

$$\text{cov}_{(x,y)} = (1/2) \sigma_a^2 + (1/4) \sigma_d^2 + (1/4) \sigma_{aa}^2 + (1/8) \sigma_{ad}^2 + (1/16) \sigma_{dd}^2 + (1/8) \sigma_{aaa}^2$$

(c) Parentesco entre mãe e filho(a) ou entre pai e filho(a):

$$\alpha = 1/2; \quad \delta = 0; \quad \alpha^2 = 1/4; \quad \alpha\delta = 0; \quad \delta^2 = 0; \quad \alpha^3 = 1/8 \quad (6)$$

$$\text{cov}_{(x,y)} = (1/2) \sigma_a^2 + (1/4) \sigma_{aa}^2 + (1/8) \sigma_{aaa}^2$$

(d) Parentesco entre avô e neto:

$$\alpha = 1/4; \quad \delta = 0; \quad \alpha^2 = 1/16; \quad \alpha\delta = 0; \quad \delta^2 = 0; \quad \alpha^3 = 1/64 \quad (7)$$

$$\text{cov}_{(x,y)} = (1/4) \sigma_a^2 + (1/16) \sigma_{aa}^2 + (1/64) \sigma_{aaa}^2$$

Verifica-se que o parentesco entre avô e neto equivale exatamente ao parentesco entre meios-irmãos.

(e) Parentesco entre tio e sobrinho:

$$\alpha = 1/4; \quad \delta = 0; \quad \alpha^2 = 1/16; \quad \alpha\delta = 0; \quad \delta^2 = 0; \quad \alpha^3 = 1/64 \quad (8)$$

$$\text{cov}_{(x,y)} = (1/4) \sigma_a^2 + (1/16) \sigma_{aa}^2 + (1/64) \sigma_{aaa}^2$$

Verifica-se que o parentesco entre tio e sobrinho equivale exatamente ao parentesco entre meios-irmãos e ao parentesco entre avô e neto.

(f) Parentesco entre meio-tio e sobrinho:

$$\alpha = 1/8; \quad \delta = 0; \quad \alpha^2 = 1/64; \quad \alpha\delta = 0; \quad \delta^2 = 0; \quad \alpha^3 = 1/256 \quad (9)$$

$$\text{cov}_{(x,y)} = (1/8) \sigma_a^2 + (1/64) \sigma_{aa}^2 + (1/256) \sigma_{aaa}^2$$

A variação genética dentro (σ_{gd}^2) de grupos de indivíduos (ou seja, dentro de famílias) é dada por:

$$\sigma_{gd}^2 = (1-\alpha) \sigma_a^2 + (1-\delta) \sigma_d^2 + (1-\alpha^2) \sigma_{aa}^2 + (1-\alpha\delta) \sigma_{ad}^2 + (1-\delta^2) \sigma_{dd}^2 + (1-\alpha^3) \sigma_{aaa}^2.$$

Tem-se então, a seguinte partição da variação genética entre (σ_g^2) e dentro (σ_{gd}^2) de famílias de acordo com o tipo de família:

(g) Famílias de meios-irmãos:

$$\sigma_g^2 = \text{cov}_{(x,y)} = (1/4) \sigma_a^2 + (1/16) \sigma_{aa}^2 + (1/64) \sigma_{aaa}^2$$

$$\sigma_{gd}^2 = (3/4) \sigma_a^2 + \sigma_d^2 + (15/16) \sigma_{aa}^2 + \sigma_{ad}^2 + \sigma_{dd}^2 + (63/64) \sigma_{aaa}^2$$

(h) Famílias de irmãos germanos:

$$\sigma_g^2 = \text{cov}_{(x,y)} = (1/2) \sigma_a^2 + (1/4) \sigma_d^2 + (1/4) \sigma_{aa}^2 + (1/8) \sigma_{ad}^2 + (1/16) \sigma_{dd}^2 + (1/8) \sigma_{aaa}^2$$

$$\sigma_{gd}^2 = (1/2) \sigma_a^2 + (3/4) \sigma_d^2 + (3/4) \sigma_{aa}^2 + (7/8) \sigma_{ad}^2 + (15/16) \sigma_{dd}^2 + (7/8) \sigma_{aaa}^2$$

(i) Famílias de irmãos germanos (IG) e de meios-irmãos (MI) obtidas sob cruzamentos hierárquicos:

$$\sigma_{g_{MI}}^2 = (1/4) \sigma_a^2 + (1/16) \sigma_{aa}^2 + (1/64) \sigma_{aaa}^2$$

$$\sigma_{g_{IG-MI}}^2 = (1/4) \sigma_a^2 + (1/4) \sigma_d^2 + (3/16) \sigma_{aa}^2 + (1/8) \sigma_{ad}^2 + (1/16) \sigma_{dd}^2 + (7/64) \sigma_{aaa}^2 = \text{COV}_{(x,y)_G} - \text{COV}_{(x,y)_M}$$

$$\sigma_{gd_{IG}}^2 = (1/2) \sigma_a^2 + (3/4) \sigma_d^2 + (3/4) \sigma_{aa}^2 + (7/8) \sigma_{ad}^2 + (15/16) \sigma_{dd}^2 + (7/8) \sigma_{aaa}^2$$

(j) Famílias de irmãos germanos (IG) e de meios-irmãos maternos (MIM) e paternos (MIP) obtidas sob cruzamentos fatoriais ou dialélicos:

$$\sigma_{g_{MIM}}^2 = (1/4) \sigma_a^2 + (1/16) \sigma_{aa}^2 + (1/64) \sigma_{aaa}^2$$

$$\sigma_{g_{MIP}}^2 = (1/4) \sigma_a^2 + (1/16) \sigma_{aa}^2 + (1/64) \sigma_{aaa}^2$$

$$\sigma_{g_{IG-MXP}}^2 = \text{cov}_{(x,y)_G} - 2\text{cov}_{(x,y)_M} = (1/4) \sigma_d^2 + (1/8) \sigma_{aa}^2 + (1/8) \sigma_{ad}^2 + (1/16) \sigma_{dd}^2 + (6/64) \sigma_{aaa}^2$$

Com base nas variâncias e covariâncias apresentadas, verifica-se:

- Os coeficientes ($\alpha^2, \alpha^3, \dots, \delta^2, \delta^3, \dots$) das variâncias epistáticas tendem a diminuir com o aumento do grau de interação, de forma que os efeitos de maior ordem que aaa podem ser negligenciados nas expressões das covariâncias entre parentes.
- Ignorando-se os efeitos epistáticos, a variância genética aditiva (σ_a^2) pode ser estimada a partir da avaliação de progênies de meios-irmãos ou do uso de delineamentos de cruzamento hierárquico ou fatorial ou dialélico.
- Ignorando-se os efeitos epistáticos, a variância genética de dominância (σ_d^2) pode ser estimada a partir do uso dos delineamentos de cruzamento hierárquico ou fatorial ou dialélico, os quais baseiam-se, simultaneamente, na avaliação de progênies de meios-irmãos e irmãos germanos. O uso exclusivo de progênies de irmãos germanos não permite separar (σ_a^2) e (σ_d^2), e, portanto, não permite estimar estes parâmetros.

Existem evidências de que as variâncias epistáticas apresentam menores magnitudes do que (σ_a^2) e (σ_d^2) em populações panmíticas (Vencovsky, 1969). Isto tem conduzido à suposição de variâncias epistáticas nulas, para efeito de uso da covariância entre parentes na estimação de (σ_a^2) e (σ_d^2).

A estimação da variância epistática aditiva x aditiva pode ser realizada em espécies panmíticas perenes, desde que sejam incluídos na análise dados de genitores e de suas progênies, e, portanto, existam relações de parentesco mãe-filho(a) ou pai-filho(a) no conjunto de indivíduos analisados. Assim, as seguintes variâncias podem ser estimadas:

- (σ_a^2) e (σ_{aa}^2), a partir de conjuntos de dados envolvendo relações de parentesco de meios-irmãos, entre pais e filhos e ignorando-se as frações (1/64) σ_{aaa}^2 e (1/8) σ_{aaa}^2 em (4) e (6), respectivamente.
- (σ_a^2), (σ_d^2) e (σ_{aa}^2), a partir de conjuntos de dados que englobem as relações de parentesco de meios-irmãos, irmãos germanos e pai-filho, e ignorando as frações (1/64) σ_{aaa}^2 e (1/8) σ_{aaa}^2 em (4) e (6) e (1/8) $\sigma_{aaa}^2 + (1/8) \sigma_{ad}^2 + (1/16) \sigma_{dd}^2$ em (5).

Para a estimação de outros componentes da variância epistática são necessárias mais de três relações de parentesco entre os indivíduos no conjunto de dados analisados. As demais relações de parentesco (avô-neto, tio-sobrinho e meio-tio-sobrinho) apresentam baixas frações da variância σ_{aaa}^2 e, portanto, dificilmente contribuirão efetivamente para estimação de um quarto componente de variância genética. Na estimação dos três componentes mencionados σ_a^2 , σ_d^2 e σ_{aa}^2 , a maioria da informação para σ_{aa}^2 advém das relações de parentesco entre pai-filho e entre irmãos germanos, ao passo que para σ_d^2 toda a infor-

mação é propiciada pela relação de parentesco de irmãos germanos e para σ_a^2 as informações provêm de todas as relações de parentesco mencionadas.

Avaliando-se, também, indivíduos clonados, tem-se, ainda, a seguinte relação de parentesco entre dois indivíduos x e y , clonados:

$$\text{cov}_{(x,y)} = \sigma_a^2 + \sigma_d^2 + \sigma_{aa}^2 + \sigma_{ad}^2 + \sigma_{dd}^2 + \sigma_{aaa}^2$$

De posse de mais esta relação de parentesco, adicionalmente aos parentescos de pai-filho, irmãos germanos e meios-irmãos, torna-se possível estimar um quarto componente de variância, que deve ser σ_{aaa}^2 , pois está presente em todas as quatro relações de parentesco, ou mesmo uma função linear de variâncias epistáticas contendo σ_{ad}^2 , σ_{dd}^2 e σ_{aaa}^2 , função esta dada por $\sigma_{ad}^2 + \sigma_{dd}^2 + \sigma_{aaa}^2 + \dots$.

9.14.3. Modelo linear misto incluindo efeitos epistáticos

Em notação matricial, o modelo para os valores fenotípicos de um caráter quantitativo em uma população panmítica pode ser expresso por:

$$y = Xb + Za + Zd + Zaa + Zad + Zdd + Zaaa + e \quad (10), \text{ em que:}$$

X e Z = matrizes de incidência para os vetores de efeitos fixos (b) e para os vetores aleatórios de efeitos genéticos (a , d , aa , ad , dd e aaa), respectivamente.

O vetor b contempla os efeitos ambientais identificáveis tais como o efeito de bloco.

A esperança e a variância dos vetores aleatórios são definidas como:

$$E \begin{bmatrix} y \\ a \\ d \\ aa \\ ad \\ dd \\ aaa \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \text{Var} \begin{bmatrix} y \\ a \\ d \\ aa \\ ad \\ dd \\ aaa \\ e \end{bmatrix} = \begin{bmatrix} V & ZG_1 & ZG_2 & ZG_3 & ZG_4 & ZG_5 & ZG_6 & R \\ G_1Z & G_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ G_2Z & 0 & G_2 & 0 & 0 & 0 & 0 & 0 \\ G_3Z & 0 & 0 & G_3 & 0 & 0 & 0 & 0 \\ G_4Z & 0 & 0 & 0 & G_4 & 0 & 0 & 0 \\ G_5Z & 0 & 0 & 0 & 0 & G_5 & 0 & 0 \\ G_6Z & 0 & 0 & 0 & 0 & 0 & G_6 & 0 \\ R & 0 & 0 & 0 & 0 & 0 & 0 & R \end{bmatrix}, \text{ em que:}$$

$$G_1 = A\sigma_a^2; \quad G_4 = A\#D\sigma_{ad}^2$$

$$G_2 = D\sigma_d^2; \quad G_5 = D\#D\sigma_{dd}^2$$

$$G_3 = A\#A\sigma_{aa}^2; \quad G_6 = A\#A\#A\sigma_{aaa}^2$$

$$R = I\sigma_e^2: \text{ matriz de covariância residual.}$$

$$V = ZAZ\sigma_a^2 + ZDZ\sigma_d^2 + ZA\#AZ\sigma_{aa}^2 + ZA\#DZ\sigma_{ad}^2 + ZD\#DZ\sigma_{dd}^2 + ZA\#A\#AZ\sigma_{aaa}^2 + I\sigma_e^2 = \text{matriz de covariância fenotípica.}$$

G_i : matriz de covariância genética para o efeito i .

I : matriz identidade.

A : matriz de parentesco (correlação) genético aditivo.

D : matriz de parentesco (correlação) genético de dominância.

$A\#A$: matriz de parentesco (correlação) genético epistático do tipo aditivo x aditivo.

$A\#D$: matriz de parentesco (correlação) genético epistático do tipo aditivo x dominância.

$D\#D$: matriz de parentesco (correlação) genético epistático do tipo dominância x dominância.

$A\#A\#A$: matriz de parentesco (correlação) genético epistático do tipo aditivo x aditivo x aditivo.

: operação produto de Hadamard entre duas matrizes, significando que, se $M = A\#A$, então

$m_{ij} = (a_{ij})(a_{ij}) = a_{ij}^2$ é o produto direto dos elementos de A .

9.14.4. Melhor predição linear não viciada (BLUP) dos efeitos epistáticos

A metodologia denominada melhor predição linear não viciada (BLUP) foi desenvolvida por Henderson (1949) para a predição dos efeitos genéticos e estimação dos efeitos ambientais, simultaneamente, a qual apresenta propriedades estatísticas e práticas ótimas. Associado a essa metodologia, Henderson (1950) apresentou um sistema de equações lineares, denominado equações de modelo misto (MME) (publicadas em detalhe por Henderson, 1973), o qual permite a estimação dos efeitos fixos e a predição de efeitos genéticos sem a necessidade de inversão da matriz V . Considerando conhecidos os componentes de variância, as MME para a predição dos efeitos genéticos sob o modelo (10), ignorando os efeitos ad , dd e aaa , equivalem a:

$$\begin{bmatrix} X'X & X'Z & X'Z & X'Z \\ ZX & ZZ + A^{-1}\lambda_1 & ZZ & ZZ \\ ZX & ZZ & ZZ + D^{-1}\lambda_2 & ZZ \\ ZX & ZZ & ZZ & ZZ + (A\#A)^{-1}\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{d} \\ \hat{aa} \end{bmatrix} = \begin{bmatrix} X'y \\ Zy \\ Zy \\ Zy \end{bmatrix} \text{ ou } C \hat{s} = v \quad (11), \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2}; \quad \lambda_2 = \frac{\sigma_e^2}{\sigma_d^2}; \quad \lambda_3 = \frac{\sigma_e^2}{\sigma_{aa}^2}.$$

C : matriz dos coeficientes.

\hat{s} : vetor de solução.

v : vetor a direita associado às equações de modelo misto.

Henderson (1976), após cerca de 20 anos de pesquisas, desenvolveu um eficiente algoritmo que permite escrever A^{-1} diretamente a partir de uma lista de genitores (pedigree) de forma que a matriz de parentesco aditivo não precisa ser invertida. Também D^{-1} (Hoeschele & Van Raden, 1991) e $(A\#A^{-1})$ (Van Ra-

den & Hoeschele, 1991) podem ser escritas diretamente. Dessa forma, a predição dos efeitos genéticos pode ser obtida resolvendo-se iterativamente as MME e usando uma combinação da iteração de Gauss-Seidel e Jacobi (Golub & Van Loan, 1983; Misztal & Gianola, 1987), sem a necessidade de inversão da matriz dos coeficientes das equações de modelo misto.

Uma vez que os componentes de variância, não são, na prática, conhecidos, eles são estimados e depois utilizados nas MME. Dependendo da forma como os componentes de variância são estimados, existe semelhança formal com a estimativa bayesiana empírica (Harville, 1977) e este procedimento em duas etapas é denominado BLUP empírico (Harville & Carriquiry, 1992).

9.14.5. Estimação dos componentes de variância por REML

Com o advento do modelo animal ou BLUP no modelo individual, houve uma grande mudança no pensamento sobre a estimativa de componentes de variância. Anteriormente ao modelo animal, as covariâncias entre parentes eram estimadas e interpretadas em termos de suas esperanças matemáticas (igualando-as aos seus valores esperados), gerando os componentes de variância. Hoje os componentes de variância são estimados diretamente como as variâncias dos efeitos aleatórios do modelo linear misto.

O termo modelo animal foi criado por Quaas & Pollak (1980) e o método da máxima verossimilhança restrita (REML) sob modelo animal tornou-se o método padrão para a estimação de componentes de variância e parâmetros genéticos a partir de dados desbalanceados devido às suas propriedades estatísticas superiores ao método de quadrados mínimos e de máxima verossimilhança. O método REML foi desenvolvido por Patterson & Thompson (1971) e Thompson (1973).

Atribuindo valores iniciais para os componentes de variância na expressão (11), são obtidas predições para os efeitos genéticos a , d e aa . Calculando as variâncias desses efeitos preditos, resultam as variâncias σ_a^2 (variância genética aditiva), σ_d^2 (variância genética de dominância) e σ_{aa}^2 (variância epistática do tipo aditivo x aditivo), as quais, provavelmente, serão diferentes dos valores iniciais utilizados na expressão (11), significando que os valores iniciais utilizados não foram plausíveis ou verossímeis. Dessa forma, as MME devem ser resolvidas novamente usando os componentes de variância calculados com base nas predições dos efeitos aleatórios obtidos dos dados experimentais. Procedendo-se, sucessivamente, dessa maneira, atinge-se a convergência para os componentes de variância, ou seja, os valores utilizados nas MME equivalem às próprias variâncias dos efeitos preditos, significando que os valores utilizados nas MME passaram a ser plausíveis ou verossímeis com o conjunto de dados. Este é o princípio da verossimilhança.

Os estimadores REML para iteração dos componentes de variância na expressão (11) empregando-se o algoritmo EM (Expectation – Maximization) de Dempster et al. (1977) são:

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

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

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

$$\hat{\sigma}_{aa}^2 = [\hat{aa}(A\#A)^{-1}\hat{aa} + \hat{\sigma}_e^2 \operatorname{tr} (A\#A)^{-1} C^{44}]/q, \text{ em que:}$$

tr: operador traço matricial.

r(x): posto da matriz *X*.

N-r(x): graus de liberdade do erro.

q: número de indivíduos.

N: número de dados.

*C*²², *C*³³ e *C*⁴⁴ provêm de:

$$C = \begin{bmatrix} C^{11} & C^{12} & C^{13} & C^{14} \\ C^{21} & C^{22} & C^{23} & C^{24} \\ C^{31} & C^{32} & C^{33} & C^{34} \\ C^{41} & C^{42} & C^{43} & C^{44} \end{bmatrix} : \quad \begin{array}{l} \text{inversa generalizada da matriz dos coeficientes} \\ \text{das MME.} \end{array}$$

Verifica-se que, em associação com o modelo misto individual (modelo animal), o método REML estima os componentes de variância diretamente como as variâncias dos efeitos aleatórios do modelo, mas considera também todos os parentescos entre os indivíduos em avaliação. Na estimação são considerados, ainda, os efeitos ambientais sistemáticos ou fixos do modelo, de forma que os componentes de variância são válidos dentro de (e também entre) cada nível dos efeitos fixos. Em resumo, a iteração dos componentes de variância (variância dos efeitos aleatórios preditos) nas equações de modelo misto permite verificar a verossimilhança de particulares valores numéricos dos parâmetros, ou seja, possibilita constatar a plausibilidade de que os dados tenham sido amostrados de uma população com os parâmetros avaliados.

9.14.6. Importância do ajuste dos efeitos de dominância e dos efeitos epistáticos do tipo aditivo x aditivo para a estimação de parâmetros genéticos e predição dos efeitos genéticos

A importância dos ajustes dos efeitos de dominância e epistáticos para o melhoramento de espécies perenes pode ser caracterizada considerando as duas estruturas de dados mais comuns no melhoramento dessas espécies.

- (1) Para conjuntos de dados que envolvem relações de parentesco de meios-irmãos e de pais-filhos, o ajuste do efeito da epistasia do tipo aditivo x aditivo permite melhorar, sensivelmente, os modelos, conduzindo a estimativas mais precisas da herdabilidade.

Neste caso, os efeitos na estimativa da herdabilidade (h_a^2) são:

- (i) Eliminação de $4(1/16) \sigma_{aa}^2 = 1/4 \sigma_{aa}^2$ da contribuição da covariância de meios-irmãos para o numerador da h_a^2 .
 - (ii) Eliminação de $2(1/4) \sigma_{aa}^2 = 1/2 \sigma_{aa}^2$ da contribuição da covariância de pais e filhos para o numerador da h_a^2 .
- (2) Para conjuntos de dados que envolvem relações de parentesco de meios-irmãos, irmãos germanos e pais e filhos, o ajuste do efeito de dominância e da epistasia do tipo aditivo x aditivo conduz aos seguintes efeitos na estimativa da herdabilidade:
- (i) Eliminação de $4(1/16) \sigma_{aa}^2 = 1/4 \sigma_{aa}^2$ da contribuição da covariância de meios-irmãos para o numerador da h_a^2 .
 - (ii) Eliminação de $2(1/4) \sigma_{aa}^2 = 1/2 \sigma_{aa}^2$ da contribuição da covariância de pais e filhos para o numerador da h_a^2 .
 - (iii) Eliminação de $2(1/4) \sigma_{aa}^2 = 1/2 \sigma_{aa}^2$ da contribuição da covariância de irmãos germanos para o numerador da h_a^2 .
 - (iv) Eliminação de $2(1/4) \sigma_d^2 = 1/2 \sigma_d^2$ da contribuição da covariância de irmãos germanos para o numerador da h_a^2 .

Verifica-se assim que, se σ_d^2 e σ_{aa}^2 apresentam magnitudes consideráveis, o não ajuste de d e aa provoca uma grande superestimação da herdabilidade no sentido restrito (h_a^2). Adicionalmente à maior precisão na estimativa da h_a^2 , os próprios \hat{d} e \hat{aa} ajustados são úteis na seleção de indivíduos para o mérito genotípico total e não apenas aditivo.

Estruturas de dados do tipo (1) são muito comuns no melhoramento de espécies vegetais perenes quando se utilizam dados de duas gerações consecutivas. Estruturas de dados do tipo (2) são também comuns no melhoramento de espécies vegetais perenes quando se utilizam os delineamentos de cruzamentos fatoriais (ou dialélicos ou hierárquicos) e dados de gerações consecutivas. Tanto na situação (1) quanto na situação (2), o ajuste de aa afeta também as contribuições das covariâncias avô-neto, tio-sobrinho e meio-tio-sobrinho para o numerador da h_a^2 , em casos de dados de muitas gerações serem utilizados. Neste caso, elimina-se $(1/4) \sigma_{aa}^2$ da contribuição das expressões (7) e (8) para h_a^2 e $(1/8) \sigma_{aa}^2$ da contribuição da expressão (9) para a superestimação da h_a^2 .

Assim, o ajuste de aa deve, sempre que possível, ser realizado e o ajuste de d deve ser realizado em conjuntos de dados que apresentam grande número de relações de parentesco de dominância.

Quanto aos efeitos de dominância, uma maneira de eliminar suas influências na estimação da h^2 é ajustar o efeito da família de irmãos germanos ou interação pai x mãe. Este ajuste remove da contribuição da covariância de irmãos germanos para o numerador da h^2 , a fração

$$\begin{aligned} & 2[(1/4) \sigma_d^2 + (1/8) \sigma_{aa}^2 + (1/8) \sigma_{ad}^2 + (1/16) \sigma_{dd}^2 + (6/64) \sigma_{aaa}^2] = \\ & = (1/2) \sigma_d^2 + (1/4) \sigma_{aa}^2 + (1/4) \sigma_{ad}^2 + (1/8) \sigma_{dd}^2 + (3/16) \sigma_{aaa}^2. \end{aligned}$$

Embora este artifício represente uma grande economia computacional, não permite estimar o efeito individual d, o qual pode ser útil na seleção para o mérito genotípico.

9.15. Vantagens do Uso da Metodologia de Modelos Mistos para a Estimação de Componentes de Variância e Predição de Valores Genéticos

A metodologia de modelos mistos é uma ferramenta flexível para a estimação de parâmetros genéticos e predição de valores genéticos e apresenta as seguintes vantagens:

- Pode ser aplicada a dados desbalanceados.
- Permite utilizar simultaneamente um grande número de informações, provenientes de diferentes gerações, locais e idades, gerando estimativas e predições mais precisas.
- Não exige dados obtidos sob estruturas rígidas de experimentação, podendo ser aplicada a dados obtidos normalmente nos programas de melhoramento, os quais não precisam estar associados a delineamentos, bastando que se tenham informações sobre a genealogia dos indivíduos.
- Permite a estimação dos efeitos de dominância e epistáticos, além dos aditivos, pois utiliza maior número de relações de parentesco.
- Permite o desenvolvimento de programas nacionais de avaliação genética, fato que permite elevar a taxa global de melhoramento da espécie.
- Permite o ajuste de vários modelos alternativos, podendo-se escolher o que se ajusta melhor aos dados e, ao mesmo tempo, é parcimonioso (apresenta menor número de parâmetros).
- Permite comparar indivíduos através do tempo e do espaço.
- Permite lidar com estruturas complexas de dados (medidas repetidas, diferentes anos, locais e delineamentos).
- Permite a simultânea correção para os efeitos ambientais, estimação de componentes de variância e predição de valores genéticos.

Em função dessas vantagens, a estimativa de parâmetros genéticos e predição de valores genéticos devem basear-se no procedimento REML/BLUP sob modelo individual, para todas as espécies perenes em que os dados são obtidos ao nível de indivíduo. As propriedades teóricas desejáveis da metodologia de modelos mistos com vistas à estimativa de componentes de variância são apresentadas com mais detalhes por Robinson (1987); Searle et al. (1992), Rao (1997) e Lynch & Walsh (1997).

9.16. 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 & Wedderburn (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.

Nessa situação, a variável de trabalho pode ser definida pelo modelo $y = \mu + (y - \mu)g'(\mu)$, em que μ refere-se à esperança de y e $(y - \mu)g'(\mu)$ 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,

$$Cov(y) = ZGZ + R = ZGZ + I\sigma_e^2.$$

Quando Y segue uma distribuição binomial, têm-se a estimativa e a 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:

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

De posse da variável observacional (ou dependente) ajustada y^* , o modelo linear misto equivale a $y^* = Xb + Za + (y - \mu)g'(\mu)$, em que:

$$E(y^*) = Xb, \quad Cov(a) = G, \quad Cov[(y - \mu)g'(\mu)] = W^{-1}\sigma_e^2 \quad e \quad Cov(y^*) = ZGZ + W^{-1}\sigma_e^2.$$

O modelo $y^* = Xb + Za + (y - \mu)g'(\mu)$ apresenta 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 $Cov[(y - \mu)g'(\mu)] = W^{-1}\sigma_e^2$.

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

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

$$S^{-1} = \text{matriz com termos diagonais dados por } \mu_i (1 - \mu_i) \frac{1}{\sigma_{e_L}^2};$$

$\sigma_{e_L}^2$ = variância residual na escala contínua de tolerância (*liability*).

b_L e a_L = efeitos fixos e aleatórios na escala de tolerância.

Quando a refere-se a um vetor de valores genéticos aditivos, $Cov(a) = G = A\sigma_{a_L}^2$, 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:

- estimação de $\mu = n_1/N$, em que n_1 é o número de indivíduos que recebe o escore 1, dentre N indivíduos avaliados.
- obtenção de y^* , a partir de y e μ (nesta etapa, a variável passa do intervalo (0,1) para a reta real, ou seja, a função é linearizada).
- estimação de b_L e \hat{a}_L , dados os valores atuais ou correntes de μ , $\sigma_{e_L}^2$ e $\sigma_{a_L}^2$.
- obtenção de $\hat{\sigma}_{e_L}^2$ e $\hat{\sigma}_{a_L}^2$ iterativamente e, após a convergência, proceder a obtenção atualizada de \hat{b}_L e \hat{a}_L .
- obtenção de $\hat{\eta} = \hat{\theta} = X\hat{b}_L + Z\hat{a}_L$.

(f) obtenção de novo valor predito de μ , usando a função de ligação, através de

$$\hat{\mu}_L = \frac{e^{\hat{\theta}}}{1+e^{\hat{\theta}}} \quad (\text{neste passo, a variável volta ao intervalo } (0,1)).$$

(g) atualização de S^{-1} via $S^{-1} = \hat{\mu}_1(1-\hat{\mu}_1) \frac{1}{\hat{\sigma}_{\theta L_1}^2}$ 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** que compreende os passos de (a) a (h), a necessidade de convergência no passo (d).

Software

Dentre os softwares mencionados neste documento, o ASREML é o único adequado para a análise de variáveis binomiais. Para a versão Windows, o arquivo executável necessário é o ASRWIN.EXE. A utilização do referido aplicativo na análise de modelos mistos ao nível de plantas individuais requer três 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 por meio do programa Notepad (ou bloco de notas, usando documento texto, ou seja extensão .txt) ou mesmo o Excel por meio da opção salvar como “Texto” (OS/2 ou MS-DOS), o que produzirá uma extensão .txt.

O arquivo de comandos deve conter a extensão .as e deve ser escrito com base em cinco 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, mediante modificação de arquivos preexistentes. Os arquivos de resultados mais importantes do programa são aqueles com extensão .sln (que apresenta a solução para todos os efeitos do modelo com seus respectivos desvios padrões) e .asr (que 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 progénies de meios-irmãos (polinização aberta) e 6 blocos, em que foi avaliada a variável sobrevivência (Sob). Tendo salvos os arquivos de dados e de pedigree com os nomes Dados.txt e Pedigree.txt, o arquivo com extensão .as é dado por:

Título

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, os comandos advêm após o símbolo!.

Assim:

P : indica que o número de indivíduos deve ser lido no arquivo de pedigree.

SKIP 1: indica que deve ser ignorada a primeira linha dos arquivos, pois se refere 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:

Título

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 de cruzamentos dialélicos com medidas repetidas em cada indivíduo, tem-se o seguinte arquivo .as, considerando cruzamentos de 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

Aqui os efeitos de bloco e medição foram ajustados como fixos. O comando REPEAT indica que se trata de um caso de medidas repetidas.

Depois de 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. Muitas vezes, as categorias surgem porque não é possível medir a variável real de interesse. Tomar as variáveis categóricas como normais é tanto mais apropriado 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 às variáveis.

Considere a avaliação da variável número de frutos (denominada Frutos) de cacauzeiros em dois locais, no delineamento em blocos ao acaso com 18 progenies 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, equivale 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 !M0 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 postos em duas colunas distintas, preenchendo os dados inexistentes com zero (neste caso).

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 existirem 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, ve₁ e ve₂ dizem respeito aos valores iniciais para as variâncias residuais nos locais 1 e 2, respectivamente, e vc₁ e vc₂ são relativos aos valores iniciais para as variâncias entre parcelas nos locais 1 e 2, respectivamente. Por sua vez, va₁, va₂ e va₁₂ 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. Assim, tais valores numéricos devem ser fornecidos na estruturação do arquivo de comandos.

9.17. Ajuste de Covariável e Análise de Covariância

O ajuste de covariáveis é importante como forma de reduzir o erro experimental, por meio 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, é a situação em que os indivíduos são avaliados em diferentes idades, para uma determinada característica. Neste caso, a idade deve ser ajustada, no modelo, como uma covariável.

Considerando um modelo estatístico tradicional e incluindo como uma covariável x , o número de plantas na parcela, tem-se que (Steel e Torrie, 1980; Vencovsky e Barriga, 1992; Ramalho et al., 2000):

$$Y_{ij} = \mu + p_i + r_j + b(x_{ij} - \bar{x}) + e_{ij}, \text{ em que:}$$

Y_{ij} : observação da variável dependente de interesse, referente à progénie i no bloco j .

μ, p_i, r_j, e_{ij} : efeitos da média geral, da progénie 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 residual 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 (progêñies). Isto implica que não pode haver diferenças significativas entre progêñies 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 ela ocorrer de forma aleatória. Para verificar esta suposição, recomenda-se a análise de variância e a verificação da significância da fonte de variação progêñies. A significância desta fonte de variação indica que existem diferenças genéticas entre progêñies para o caráter e, portanto, o ajuste do número de plantas como covariável não é recomendado. Então, é melhor realizar a predição de valores genéticos para a sobrevivência e, posteriormente, utilizá-la também na seleção, conforme Resende (1999b).

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 diferen-

ças de sobrevivência entre progêneres signifique que tal variável não está afetando suas comparações 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 à predição, ou seja, os valores fenotípicos em y são corrigidos. Assim, os modelos, os estimadores e os 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. Uma análise suficiente para contornar tal efeito é a consideração como covariável, do número de plantas vizinhas presentes ao redor de cada planta, dentre um número máximo de 8 (número de plantas que circundam cada planta, quando todas as vizinhas sobrevivem).

Em resumo, a questão da sobrevivência incompleta no experimento tem três aspectos: (i) o desbalanceamento e a representação desbalanceada dos materiais genéticos em avaliação; (ii) o efeito da sobrevivência na produtividade por unidade de área, ou seja, por hectare; (iii) o beneficiamento de determinadas plantas em função da ausência de competição, devida às falhas no experimento. O aspecto relatado em (i) pode ser contornado por meio do uso do procedimento BLUP/REML. Para contornar a questão (ii) deve-se realizar a predição de valores genéticos para o caráter sobrevivência (nos casos em que a sobrevivência apresentar controle genético) e combinar os valores genéticos preditos para sobrevivência e produtividade, em um único valor genético predito para produtividade por área. O fator (iii) pode ser considerado adequadamente, corrigindo-se previamente os valores fenotípicos observados, via análise de covariância, considerando o número (no máximo 8) de plantas vizinhas como covariável.

Logicamente, esta correção associada a (iii) surtirá efeito quando houver influência de x (covariável número de plantas vizinhas) sobre y, ou seja, quando b (coeficiente de regressão residual entre x e y) for diferente de zero. Em um esquema de análise de variância, o coeficiente de regressão residual é estimado por $\hat{b} = SP_{e(x,y)} / SQ_{e(x)}$, em que $SP_{e(x,y)}$ e $SQ_{e(x)}$ dizem respeito às somas de produtos e quadrados dos resíduos (ambiente dentro da parcela) das análises de variância das variáveis (x,y) e x, respectivamente.

O modelo apresentado para y_{ij} refere-se aos dados observados, ou seja, aos dados não corrigidos para os efeitos de falha (redução na competição). Assim, para obtenção dos dados corrigidos para tais efeitos, a equação de ajuste é dada por

$$y_{ijc} = y_{ij} - \hat{b}(x_{ij} - 8), \text{ em que:}$$

y_{ijc} : dados corrigidos, prontos para integrar o vetor y na análise de modelos mistos.

y_{ij} : dados observados.

\hat{b} : estimativa do coeficiente de regressão residual entre o número de plantas vizinhas (x) e os dados do caráter avaliado (y).

x_{ij} : número de plantas vizinhas a cada planta.

8 : número máximo de plantas vizinhas.

Sendo x o número de plantas vizinhas, b tenderá a ser negativo, ou seja, quanto menor x maior será o y de cada planta. Assim, pela fórmula de y_{ijc} , verifica-se que quando $n < 8$, a correção diminuirá o valor da observação fenotípica e quando $n = 8$, nenhuma correção será aplicada a esta observação, conforme desejado.

É importante notar que \hat{b} é inerente a um coeficiente de regressão residual, devendo ser estimado livre dos efeitos de blocos e de tratamentos (progêneres ou clones). No contexto dos modelos mistos isto significa que deve-se adotar o procedimento REML/BLUP para as variáveis ($x+y$), x e y , ajustando os efeitos genéticos, de parcela (se for o caso) e de bloco. Em seguida, o coeficiente b deve ser estimado a partir das variâncias residuais de ($x+y$), x e y pela expressão $\hat{b} = \hat{Cov}_{e(x,y)} / \hat{\sigma}_{e(x)}^2$, em que:

$\hat{Cov}_{e(x,y)} = (1/2)(\hat{\sigma}_{e(x+y)}^2 - \hat{\sigma}_{e(x)}^2 - \hat{\sigma}_{e(y)}^2)$: estimativa da covariância residual (ambiental dentro de parcelas) entre x e y .

$\hat{\sigma}_{e(x)}^2$: estimativa da variância residual para a variável x .

$\hat{\sigma}_{e(y)}^2$: estimativa da variância residual para a variável y .

Pode-se também usar o modelo bivariado, obtendo a covariância ambiental entre x e y .

Obtendo-se a variável corrigida y_c , esta deve ser submetida à análise de modelos mistos, conjuntamente com uma variável binomial z associada à sobrevivência.

Capítulo 10

**O Software Selegen –
Seleção Genética Computadorizada**



10.1. Introdução

A adoção de estratégias eficientes de melhoramento genético de espécies perenes depende, sobretudo, da utilização de acurados procedimentos de previsão de valores genéticos e métodos de seleção. Assim, é recomendável a comparação de diversos métodos e procedimentos, visando à determinação do mais eficiente a cada situação experimental.

Torna-se imprescindível a disponibilidade de um software que forneça os valores genéticos preditos e compare vários métodos de seleção, em termos de acurácia, ganho genético, tamanho efetivo populacional, variância de ganhos genéticos, dentre outros. Assim, o software Selegen foi desenvolvido em 1994 (Resende et al., 1994a; Resende & Oliveira, 1997) para servir de base ao melhoramento genético de plantas perenes, podendo ser utilizado para a seleção genética de quaisquer indivíduos pertencentes a espécies florestais (eucalipto, pinus, acácia-negra, erva-mate, seringueira, dentre outras) e espécies frutíferas (acerola, cacau, café, caju, coco, cupuaçu, dendê, guaraná, maçã, pupunha, etc.). Nessas espécies, a seleção de um indivíduo é preponderante à seleção de grupos de indivíduos (exemplo: família).

Neste capítulo são apresentados aspectos referentes ao software Selegen, visando fornecer alguns aspectos de sua base estatística e informações sobre suas potencialidades.

O software Selegen foi desenvolvido para ambiente IBP/PC-DOS. É fundamentado em algoritmos que maximizam a eficiência do processo seletivo, em função de diferentes situações experimentais. Os métodos de seleção que esse software possui são recomendáveis para o melhoramento de espécies perenes, e consideram os aspectos como:

- Seleção a partir da avaliação de um pequeno número de indivíduos por família.
- Equivalência entre unidades de seleção e unidades de recombinação.
- Seleção com base em múltiplas características.
- Avaliações em várias idades.
- Quantificação do tamanho efetivo populacional.
- Relevância da restrição do incremento da endogamia nos métodos de seleção.
- Relevância do estabelecimento de distintas populações de produção de propagulos melhorados e de melhoramento.
- Diferentes herdabilidades efetivas associadas aos diferentes indivíduos.
- Diferentes quantidades e precisão das informações associadas aos candidatos à seleção.
- Primeiros momentos dos dados não estimados com precisão pelo método dos quadrados mínimos ordinários.
- Heterogeneidade associada aos segundos momentos dos dados.

O sistema está fundamentado em algoritmos que maximizam a eficiência da seleção em diferentes situações experimentais. Possui cerca de 50 métodos de seleção, univariados e multivariados, o que viabiliza, de acordo com os objetivos da seleção, um máximo aproveitamento das diversas variáveis avaliadas nos indivíduos. A seleção é baseada nos valores genético (para a propagação sexuada) e genotípico (para a propagação assexuada) preditos de cada indivíduo. Pode ser aplicado em espécies alógamas (monóicas e dióicas) ou autógamas, e com sistema reprodutivo misto, desde que se conheça a taxa de autofecundação.

10.2. Módulos Básicos

São três os módulos básicos:

1. Índice de Seleção
2. Best Linear Prediction – BLP
3. Best Linear Unbiased Prediction – (BLUP)

A expressão "Índice de Seleção", neste caso, implica utilização de um mesmo grupo de pesos (ponderadores das informações fenotípicas), de forma generalizada para todos os candidatos à seleção. Já "Best Linear Prediction" e "Best Linear Unbiased Prediction" exigem a utilização de diferentes pesos na predição dos valores genéticos associados aos diversos indivíduos. O módulo BLP refere-se a situações em que o primeiro e segundo momentos dos dados são estimados com precisão pelo método dos quadrados mínimos ordinários. Por sua vez, o módulo BLUP é inerente à situação em que os componentes de variância são considerados conhecidos, ou seja, devem ser fornecidos pelo usuário, que pode obtê-los na literatura ou em outros programas que os estimam por máxima verossimilhança restrita – REML. O módulo Selegen-REML encontra-se em implementação, tendo por base os procedimentos descritos no Capítulo 9.

10.3. Entrada de Dados e Tratamento aos Valores Zero

Os arquivos de dados a serem utilizados no Selegen devem estar no formato DBF. Algumas considerações se fazem necessárias quanto ao uso destes arquivos:

- Os nomes dos três primeiros campos (ou colunas) devem ser, obrigatoriamente: BLOCO, PROGENIE e ARVORE, tipo numérico com quatro posições.
- Os campos seguintes (variáveis utilizadas) podem conter qualquer nome válido, tipo numérico com 16 posições e até 8 decimais.
- Não existe limitação para a quantidade de variáveis utilizadas.
- Não existe limitação para a quantidade de registros utilizados.

Os registros com o valor zero podem ser considerados como significativos ou não, a critério do usuário. Na tela de entrada do arquivo de dados o usuário é questionado quanto a isto. Em caso de o zero corresponder a falha, o usuário poderá optar por ignorar as observações zeradas. Em caso de variáveis binomiais, deve-se optar pela consideração do zero.

10.4. Análises e Resultados Fornecidos pelo Módulo Índice de Seleção

1. Análise de variâncias e médias

- Quadros de análise de variância e teste F
- Coeficientes de variação e médias de blocos e progênies

2. Parâmetros genéticos e fenotípicos - análise univariada

- Componentes da variação fenotípica
- Coeficientes de herdabilidade
- Coeficientes de correlação intraclass
- Acurárias por diferentes métodos de seleção univariada

3. Seleção univariada e respostas correlacionadas

3.1. Seleção para pomar de sementes por mudas

- Individual no bloco
- Combinada no bloco
- Índice multiefeitos no bloco

3.2. Seleção para pomar de sementes clonal ou população de melhoramento

- Individual no experimento
- Combinada no experimento
- Índice multiefeitos no experimento

3.3. Seleção de genitores

3.4. Seleção de progênies

3.5. Seleção de irmãos

3.6. Seleção dentro de progênies

3.7. Seleção para Plantio Monoprogênies e Oligoprogênies (sem desbaste)

Fornece para cada procedimento de seleção univariada:

- Valor genético de cada indivíduo do experimento, na variável objetivo da seleção.
- Valor genético de cada indivíduo, nas demais variáveis analisadas.
- Ganho genético na variável objetivo da seleção.
- Ganho genético correlacionado com as demais variáveis analisadas.

- Tamanho efetivo populacional.
- Nova média do material genético após a seleção, em todas as variáveis.
- Ordem do ranking por seleção direta, em variáveis selecionadas indiretamente.

4. Análise de covariância e correlação

- Matrizes de produtos médios
- Matrizes de covariância genética
- Matrizes de covariância fenotípica
- Matrizes de correlação genética
- Matrizes de correlação fenotípica

5. Parâmetros genéticos e fenotípicos – análise multivariada

5.1. Objetivo da seleção: um caráter

- Coeficientes de predição genética associados aos vários efeitos
- Acurárias por diferentes métodos de seleção

5.2. Objetivo da seleção: agregado genotípico

- Acurárias por diferentes métodos de seleção

6. Seleção multivariada e respostas correlacionadas

6.1. Objetivo da seleção: (um caráter)

6.1.1. Seleção em pomar de sementes por mudas

- Individual no bloco
- Combinada no bloco
- Índice multiefeitos no bloco

6.1.2. Seleção em pomar de sementes clonal ou população de melhoramento

- Individual no experimento
- Combinada no experimento
- Índice multiefeitos no experimento

6.1.3. Seleção de genitores

6.1.4. Seleção de progênies

6.1.5. Seleção de irmãos

6.1.6. Seleção dentro de progênies

6.1.7. Seleção para Plantio Monoprogênes e Oligoprogênes (sem desbaste)

Fornece para cada procedimento de seleção multivariada, objetivando ganho em um só caráter:

- Valor genético de cada indivíduo do experimento, na variável objetivo da seleção.
- Valor genético de cada indivíduo, nas demais variáveis analisadas.

- Ganho genético na variável objetivo da seleção.
- Ganho genético correlacionado com as demais variáveis analisadas
- Tamanho efetivo populacional.
- Nova média do material genético após a seleção, em todas as variáveis.
- Ordem do ranking por seleção direta, nas variáveis selecionadas indiretamente.

6.2. Objetivo da seleção: (agregado genotípico)

6.2.1. Seleção em pomar de sementes por mudas

- Individual no bloco
- Combinada no bloco
- Índice multiefeitos no bloco

6.2.2. Seleção em pomar de sementes clonal, ou população de melhoramento

- Individual no experimento
- Combinada no experimento
- Índice multiefeitos no experimento

6.2.3. Seleção de genitores

6.2.4. Seleção de progênies

6.2.5. Seleção de irmãos

6.2.6. Seleção dentro de progênies

6.2.7. Seleção em plantio monoprogênies e oligoprogênies (sem desbaste)

Fornece para cada procedimento de seleção multivariada, objetivando ganho no agregado genotípico:

- Valor genético de cada indivíduo do experimento, no agregado genotípico.
- Valor genético de cada indivíduo, nas variáveis individualmente.
- Ganho genético no agregado genotípico.
- Ganho genético correlacionado em cada variável.
- Tamanho efetivo populacional.
- Nova média do material genético após a seleção, em todas as variáveis e no índice de seleção.
- Ordem do ranking, por seleção direta das variáveis selecionadas indiretamente.

7. Ottimização da seleção em função do tamanho efetivo e da endogamia

7.1. Melhor pomar biclonal

7.2. Pomar de sementes clonal

7.3. População de melhoramento

7.4. Pomar de sementes clonal - Restrição de N máx ind/fam

- 7.5. População de melhoramento - Restrição de N máx. ind./fam.
- 7.6. Pomar de sementes por mudas (restrição de 1 ind/fam/bloco)
- 7.7. Pomar de sementes testado
- 7.8. Seleção em plantios monoprogênies e oligoprogênies (com desbaste)
- 8. Seleção em testes clonais
 - 8.1. Análise de variância e médias
 - 8.2. Parâmetros genéticos e fenotípicos
 - 8.3. Seleção clonal univariada
 - 8.4. Seleção clonal multivariada
 - 8.4.1. Objetivo da seleção: um caráter
 - 8.4.2. Objetivo da seleção: agregado genotípico
 - 9. Ordenamento de valores genéticos
 - Seleção envolvendo vários experimentos em um mesmo local
 - Seleção envolvendo diferentes procedências
- 10. Intervalo de confiança/desvio padrão de ganhos genéticos
- 11. Seleção de cruzamentos/predição da descendência

Nas Tabelas 10.1, 10.2, 10.3 são apresentados exemplos de algumas saídas do Selegen, referentes à análise de um experimento de *Eucalyptus grandis*, aos três anos de idade, instalado no delineamento de blocos ao acaso com 35 progênies, seis repetições, e seis árvores por parcela.

Tabela 10.1. Resultados referentes a parâmetros genéticos e fenotípicos – análise univariada. Variável: diâmetro a altura do peito.

Parâmetros genéticos e fenotípicos	Estimativa
Variância	
01 – Dentro de parcelas	5,8672
02 – Ambiental entre parcelas	0,5187
03 – Entre progênies	0,2722
04 – Entre blocos	0,3212
05 – Genética aditiva	1,0888
06 – Fenotípica entre médias de progênies	0,5530
07 – Fenotípica	6,9793
Herdabilidade dos efeitos	
08 - De indivíduo na parcela	0,1392
09 - De progênies	0,5411
10 - De parcela	0,0963
11 - De blocos	0,0132
12 - De indivíduo no bloco	0,1635
13 - De indivíduo no experimento	0,1560

Continua...

Tabela 10.1. Continuação.

Parâmetros genéticos e fenotípicos	Estimativa
14 - De indivíduo dentro de progênies	0,1239
Coeficientes de correlação intraclasse (Sem ajuste para efeito de bloco)	
15 - Entre indiv. de uma mesma família em diferentes blocos	0,0390
16 – Entre indivíduos de uma mesma família no mesmo bloco	0,1593
17 – Entre indivíduos de diferentes famílias no mesmo bloco	0,0460
Coeficientes de correlação intraclasse (Com ajuste para efeito de bloco)	
18 - Entre indivíduos de uma mesma família em diferentes blocos	0,0409
19 – Entre indivíduos de uma mesma família no mesmo bloco	0,1188
Acurárias associadas às unidades de seleção univariadas (Pomar de sementes por mudas)	
20 – Individual no bloco	0,4044
21 – Combinada no bloco	0,4773
22 – Índice multiefeitos no bloco	0,4893
Acurárias associadas às unidades de seleção univariadas (Pomar de sementes clonal, população de melhoramento)	
23 – Individual no experimento	0,3950
24 – Combinada no experimento	0,4773
25 – Índice multiefeitos no experimento	0,4894
Acurárias associadas às unidades de seleção	
26 – Efeito de indivíduo na parcela	0,2892
27 – Efeito de progênies	0,3798
Seleção de genitores (Pomar de sementes testado)	
28 – Herdabilidade	0,9845
29 – Acurácia	0,6909
Seleção de irmãos (Recombinação via sementes remanescentes)	
30 – Herdabilidade	0,4922
31 – Acurácia	0,3454

Tabela 10.2. Resultados da seleção univariada e respostas correlacionadas – método índice multiefeitos no experimento.

Seleção para pomar de sementes clonal ou população de melhoramento.

Seleção índice multiefeitos no experimento.

Número total de registros: 1.188

Percentual selecionado: 2,53%

Número de indivíduos selecionados: 30

Ordem	Bloco	Progénie	Árvore	Valor genético	Ganho acum.	Ganho acum. (%)	Nova média	Tamanho efetivo populacional
Variável: diâmetro a altura do peito								
1	5	11	6	1,8020	1,8020	14,1319	14,5535	1,0000
2	4	42	6	1,4472	1,6246	12,7405	14,3761	2,0000

Continua...

Tabela 10.2. Continuação.

Ordem	Bloco	Progênie	Árvore	Valor genético	Ganho acum.	Ganho acum. (%)	Nova média	Tamanho efetivo populacional
Variável: diâmetro a altura do peito								
3	5	104	3	1,3598	1,5364	12,0484	14,2879	3,0000
4	2	2	2	1,3530	1,4905	11,6890	14,2420	4,0000
5	2	104	5	1,2756	1,4475	11,3518	14,1990	4,4944
6	2	91	1	1,2575	1,4159	11,1034	14,1674	5,4962
7	4	8	5	1,2500	1,3922	10,9176	14,1437	6,4972
8	3	104	3	1,2350	1,3725	10,7635	14,1240	6,6207
9	6	104	2	1,2280	1,3564	10,6376	14,1080	6,5454
10	4	91	1	1,2158	1,3424	10,5273	14,0939	7,2115
11	4	42	4	1,2036	1,3298	10,4284	14,0813	7,8870
12	4	104	5	1,1899	1,3181	10,3370	14,0696	7,7419
13	2	42	1	1,1613	1,3061	10,2424	14,0576	8,1873
14	4	45	2	1,1243	1,2931	10,1406	14,0446	9,0812
15	2	42	3	1,1168	1,2813	10,0484	14,0328	9,2966
16	3	42	4	1,0898	1,2693	9,9545	14,0209	9,3455
17	2	93	5	1,0823	1,2583	9,8683	14,0099	10,2060
18	5	42	5	1,0817	1,2485	9,7913	14,0000	10,1081
19	4	36	4	1,0707	1,2392	9,7179	13,9907	10,9509
20	2	104	6	1,0543	1,2299	9,6454	13,9814	10,9070
21	3	4	1	1,0423	1,2210	9,5753	13,9725	11,7348
22	2	42	2	1,0277	1,2122	9,5064	13,9637	11,5728
23	1	1	5	1,0264	1,2041	9,4430	13,9556	12,3822
24	3	36	5	1,0170	1,1963	9,3819	13,9478	13,0344
25	1	91	5	1,0006	1,1885	9,3205	13,9400	13,5268
26	1	91	4	0,9797	1,1805	9,2575	13,9320	13,8628
27	1	93	3	0,9794	1,1730	9,1991	13,9245	14,5166
28	2	42	4	0,9400	1,1647	9,1338	13,9162	14,2628
29	3	91	5	0,9349	1,1568	9,0717	13,9083	14,4841
30	5	8	6	0,9166	1,1488	9,0089	13,9003	15,1249
Resposta correlacionada na variável: altura								
52	5	11	6	0,9941	0,9941	5,1338	20,3583	1,0000
1	4	42	6	1,9752	1,4846	7,6670	20,8488	2,0000
109	5	104	3	0,7774	1,2489	6,4496	20,6131	3,0000
45	2	2	2	1,0246	1,1928	6,1600	20,5570	4,0000
14	2	104	5	1,2894	1,2121	6,2597	20,5763	4,4944
30	2	91	1	1,0954	1,1927	6,1592	20,5569	5,4962
8	4	8	5	1,4733	1,2328	6,3662	20,5969	6,4972
10	3	104	3	1,4297	1,2574	6,4934	20,6216	6,6207
20	6	104	2	1,1642	1,2470	6,4399	20,6112	6,5454
81	4	91	1	0,8737	1,2097	6,2471	20,5739	7,2115
6	4	42	4	1,4967	1,2358	6,3818	20,6000	7,8870
95	4	104	5	0,8196	1,2011	6,2027	20,5653	7,7419
2	2	42	1	1,8249	1,2491	6,4505	20,6133	8,1873
56	4	45	2	0,9702	1,2292	6,3476	20,5933	9,0811
26	2	42	3	1,1072	1,2210	6,3056	20,5852	9,2966
15	3	42	4	1,2876	1,2252	6,3271	20,5894	9,3455
63	2	93	5	0,9279	1,2077	6,2368	20,5719	10,2059

Continua...

Tabela 10.2. Continuação.

Ordem	Bloco	Progênie	Árvore	Valor genético	Ganho acum.	Ganho acum. (%)	Nova média	Tamanho efetivo populacional
Resposta correlacionada na variável: altura								
23	5	42	5	1,1218	1,2029	6,2122	20,5671	10,1081
208	4	36	4	0,5360	1,1678	6,0309	20,5320	10,9509
18	2	104	6	1,1698	1,1679	6,0314	20,5321	10,9070
142	3	4	1	0,6710	1,1443	5,9092	20,5084	11,7348
12	2	42	2	1,3464	1,1535	5,9567	20,5176	11,5728
51	1	1	5	0,9970	1,1467	5,9215	20,5108	12,3822
383	3	36	5	0,2516	1,1094	5,7289	20,4735	13,0344
73	1	91	5	0,8931	1,1007	5,6843	20,4649	13,5268
72	1	91	4	0,8931	1,0927	5,6435	20,4569	13,8628
49	1	93	3	1,0104	1,0897	5,6273	20,4539	14,5166
13	2	42	4	1,3464	1,0988	5,6746	20,4630	14,2628
100	3	91	5	0,8010	1,0886	5,6216	20,4528	14,4841

Tabela 10.3. Resultados da otimização da seleção em função do tamanho efetivo e da endogamia.

Otimização da seleção, em função do tamanho efetivo e da endogamia.

Pomar de sementes clonal.

Seleção univariada e respostas correlacionadas.

Seleção para pomar de sementes clonal.

Seleção índice multiefeitos no experimento

Or- dem	Blo- co	Pro- gênie	Ár- vore	Valor genético	Ganho acum.	Ganho acum. (%)	Nova média	Tamanho efetivo populacional	Ganho acumulado corrigido	Ganho acumulado corrigido (%)
Variável: diâmetro a altura do peito										
1: Indivíduos de diferentes progênies										
1	5	11	6	1,8020	1,8020	14,1318	14,5535	1,0000	0,9010	7,0660
2	4	42	6	1,4472	1,6246	12,740	14,3761	2,0000	1,2185	9,5554
3	5	104	3	1,3598	1,5364	12,048	14,2879	3,0000	1,2803	10,040
4	2	2	2	1,3530	1,4905	11,6890	14,2420	4,0000	1,3042	10,2278
5	2	91	1	1,2575	1,4439	11,3234	14,1954	5,0000	1,2995	10,1911
6	4	8	5	1,2500	1,4116	11,0699	14,1631	6,0000	1,2940	10,1474
7	4	45	2	1,1243	1,3705	10,7481	14,1221	7,0000	1,2726	9,9804
8	2	93	5	1,0823	1,3345	10,4655	14,0860	8,0000	1,2511	9,8115
9	4	36	4	1,0707	1,3052	10,2356	14,0567	9,0000	1,2327	9,6670
10	3	4	1	1,0423	1,2789	10,0295	14,0304	10,0000	1,2150	9,5280
11	1	1	5	1,0264	1,2560	9,8494	14,0075	11,0000	1,1989	9,4018
12	2	40	6	0,8053	1,2184	9,5549	13,9699	12,0000	1,1676	9,1569
13	5	41	1	0,7241	1,1804	9,2568	13,9319	13,0000	1,1350	8,9008
14	3	38	2	0,7241	1,1478	9,0012	13,8993	14,0000	1,1068	8,6797
15	2	43	2	0,7031	1,1181	8,7687	13,8697	15,0000	1,0809	8,4764

Continua...

Tabela 10.3. Continuação.

Or- dem	Blo- co	Pro- gênie	Ár- vore	Valor genético	Ganho acum.	Ganho acum. (%)	Nova média	Tamanho efetivo populacional	Ganho acumulado corrigido	Ganho acumu- lado corrigido (%)
Variável: diâmetro a altura do peito										
2: Ordenamento sem restrição										
1	5	11	6	1,8020	1,8020	14,1319	14,5535	1,0000	0,9010	7,0658
2	4	42	6	1,4472	1,6246	12,7405	14,3761	2,0000	1,2185	9,5557
3	5	104	3	1,3598	1,5364	12,0484	14,2879	3,0000	1,2803	10,0404
4	2	2	2	1,3530	1,4905	11,6890	14,2420	4,0000	1,3042	10,2278
5	2	104	5	1,2756	1,4475	11,3518	14,1990	4,4944	1,2865	10,0890
6	2	91	1	1,2575	1,4159	11,1034	14,1674	5,4962	1,2870	10,0929
7	4	8	5	1,2500	1,3922	10,9176	14,1437	6,4972	1,2850	10,0772
8	3	104	3	1,2350	1,3725	10,7635	14,1240	6,6207	1,2689	9,9510
9	6	104	2	1,2280	1,3564	10,6376	14,1080	6,5454	1,2528	9,8247
10	4	91	1	1,2158	1,3424	10,5273	14,0939	7,2115	1,2493	9,7973
11	4	42	4	1,2036	1,3298	10,4284	14,0813	7,8870	1,2455	9,7675
12	4	104	5	1,1899	1,3181	10,3370	14,0696	7,7419	1,2330	9,6695
13	2	42	1	1,1613	1,3061	10,2424	14,0576	8,1873	1,2263	9,6169
14	4	45	2	1,1243	1,2931	10,1406	14,0446	9,0811	1,2219	9,5824
15	2	42	3	1,1168	1,2813	10,0484	14,0328	9,2966	1,2124	9,5079
16	3	42	4	1,0898	1,2693	9,9545	9,9545	9,3455	1,2014	9,4216
17	2	93	5	1,0823	1,2583	9,8683	9,8683	10,2059	1,1967	9,3848
18	5	42	5	1,0817	1,0817	9,7913	9,7913	10,1081	1,1868	9,3071
19	4	36	4	1,0707	1,2392	9,7179	9,7179	10,9509	1,1826	9,2742
20	2	104	6	1,0543	1,2299	9,6454	9,6454	10,9070	1,1735	9,2028
21	3	4	1	1,0423	1,2210	9,5753	9,5753	11,7348	1,1690	9,1675
22	2	42	2	1,0277	1,2122	9,5064	9,5064	11,5728	1,1598	9,0954
23	1	1	5	1,0264	1,2041	9,4430	9,4430	12,3822	1,1555	9,0617
24	3	36	5	1,0170	1,1963	9,3819	9,3819	13,0344	1,1504	9,0217
25	1	91	5	1,0006	1,1885	9,3205	9,3205	13,5268	1,1446	8,9762
26	1	91	4	0,9797	1,1805	9,2575	9,2575	13,8628	1,1379	8,9237
27	1	93	3	0,9794	1,1730	9,1991	9,1991	14,5166	1,1326	8,8821
28	2	42	4	0,9400	1,1697	9,1338	9,1338	14,2628	1,1239	8,8139
29	3	91	5	0,9349	1,1568	9,0717	9,0717	14,4841	1,1168	8,7582
30	5	8	6	0,9166	1,1488	9,0089	9,0089	15,1249	1,1108	8,7111
Seleção ótima										
1	5	11	6	1,8020	1,8020	14,1319	14,5535	1,0000	0,9010	7,0659
2	4	42	6	1,4472	1,6246	12,7405	14,3761	2,0000	1,2185	9,5554
3	5	104	3	1,3598	1,5364	12,0484	14,2879	3,0000	1,2803	10,0403
4	2	2	2	1,3530	1,4905	11,6890	14,2420	4,0000	1,3042	10,2279

A seleção ótima apresentada (Tabela 10.3) referiu-se apenas ao ganho genético máximo corrigido para a endogamia, não considerando a variância do ganho genético estimado e o limite inferior do intervalo de confiança do ganho genético corrigido para a endogamia. Considerando as duas estatísticas mencionadas, o número ideal de indivíduos a ser selecionado situa-se ao redor de 15. Tal resultado pode ser obtido utilizando-se outro tópico do Selegen.

O Selegen permite o delineamento de eficientes programas de melhoramento para a obtenção de máximos progressos genéticos imediatos, porém compatíveis com a manutenção de variabilidade genética suficiente para o melhoramento no longo prazo. Tal aplicativo é também extremamente útil e didático para fins de ensino da Genética Quantitativa em cursos de graduação e pós-graduação.

10.5. Módulo BLP

No módulo Índice de Seleção, as quantidades usadas nos preditores de valores genéticos são: b blocos, n (média harmônica do número de plantas por parcela) plantas por parcela e nb plantas por progênies. Isto para todos os indivíduos do arquivo. O módulo BLP produz pesos específicos a cada indivíduo do experimento, considerando a sobrevivência particular de cada progênies e de cada parcela. Tal procedimento usa as seguintes quantidades: b_i blocos em que está presente a progênies i , n_{ij} plantas na parcela ij e $\sum n_i$ plantas na progênies i .

Adicionalmente ao módulo Índice de Seleção, o BLP apresenta a “herdabilidade” e a variância da média de cada progênies.

10.6. Módulo BLUP

O módulo BLUP trabalha com componentes de variância (parâmetros genéticos) estimados à parte. Para tanto é necessário informar ao programa os seguintes parâmetros: correlação genética intraclasse, herdabilidade individual no sentido restrito, herdabilidade individual no sentido amplo, repetibilidade individual e correlação devida ao ambiente comum da parcela.

Tal módulo produz pesos específicos a cada indivíduo à semelhança do BLP e prediz, também, os valores genotípicos de cada indivíduo com vistas à propagação vegetativa. Apresenta, ainda, uma decomposição completa dos valores fenotípicos (y) em efeito aditivo (a), efeito de dominância (d), efeito ambiental de parcela (c) e efeito ambiental dentro de parcela (e). Tal decomposição é muito útil tanto do ponto de vista prático quanto didático. Por exemplo, o programa decompõe a altura de 12 metros de uma planta de eucalipto em: média geral = 7; efeito aditivo = 1; efeito de dominância = 0,5; efeito de parcela = 0,5; efeito ambiental dentro de parcela = 3. Neste caso, portanto, $y = u + a + d + c + e = 7 + 1 + 0,5 + 0,5 + 3 = 12$.

10.7 Modulo REML/BLUP: Modelos Computacionais

Uma das grandes dificuldades na implementação prática da metodologia de modelos mistos ao nível individual refere-se à elevada ordem das matrizes a serem invertidas ou à resolução de sistemas de equações lineares de elevada ordem. Este tópico apresenta modelos computacionais alternativos visando contornar essas dificuldades. Tais modelos foram empregados no software Selegen-REML.

10.7.1. Modelos equivalentes

Considerando um vetor y de observações individuais, os seguintes modelos estatísticos equivalentes podem ser especificados:

- (1) $y = Xb + e_1$: modelo com interesse apenas nos efeitos fixos (MIEF).

- (2) $y = Xb + Za + e_2$: modelo de valores genéticos aditivos individuais ou modelo individual (MI).
- (3) $y = Xb + Z(a_p/2 + a_m/2 + a_d) + e_2$: modelo reduzido de valores genéticos aditivos ou modelo individual reduzido (MIR).
- (4) $y = Xb + Z_m(a_m/2) + e_4 = Xb + Z_m f + e_4$: modelo de genitores femininos ou modelo gamético (MG).

No modelo (1), o interesse reside apenas sobre os efeitos fixos (b) e todos os efeitos aleatórios (genético aditivo, genético de dominância, epistático e ambientais) são agrupados no resíduo aleatório e_1 . O modelo (2) é o próprio modelo de valores genéticos aditivos individuais (a) e, o resíduo e_2 contempla os efeitos aleatórios de dominância alélica, epistasia e ambientais. No modelo (3), o valor genético aditivo individual (a) é dividido em 3 partes: (i) metade do valor genético aditivo da mãe = $a_m/2$; (ii) metade do valor genético aditivo do pai = $a_p/2$; (iii) segregação mendeliana ou desvio genético em relação à média dos valores genéticos aditivos dos genitores = a_d . O modelo (4) é expresso em termos da metade do valor genético aditivo dos genitores femininos ou do efeito de famílias f , sendo que e_4 compreende o somatório de $a_p/2$, a_d e e_2 . Nestes modelos, X , Z e Z_m são matrizes de incidência para b , a e $a_m/2$, respectivamente.

O modelo de interesse prático ao melhoramento de plantas perenes refere-se ao (2) ou modelo individual (MI). Entretanto, tal modelo é o mais complexo computacionalmente, com número de equações para a igual ao número de descendentes mais o número de genitores em avaliação.

O modelo individual reduzido - MIR- produz resultados idênticos ao MI, porém com um menor esforço computacional, podendo-se trabalhar com um número de equações igual ao número de genitores, obtendo-se as previsões para a_p e a_m e, posteriormente, as previsões para a_d e, consequentemente, para a . Como em plantas perenes o número de descendentes é muito superior ao de genitores, o modelo (3) ou MIR deve ser utilizado.

Sob certas condições, o modelo (3) se assemelha ao modelo (4). Este é o caso, por exemplo, quando se avaliam progêniens de polinização aberta, em que os genitores masculinos são desconhecidos. Neste caso, o próprio modelo reduzido propicia previsões associadas apenas ao efeito de genitor feminino, ou seja, fornece uma medida do valor de um gameta médio desse genitor, tal como ocorre pelo modelo gamético. Da mesma forma, posteriormente, obtém-se a_d e consequentemente a .

Serão apresentados a seguir exemplos da aplicação do MI, MIR e MG associados a dados de testes de progêniens quando se conhecem ambos os genitores (testes de progêniens de irmãos germanos, por exemplo), apenas um genitor (progêniens de polinização aberta ou de meios-irmãos) e testes clonais.

10.7.2. Modelos computacionais quando se conhecem ambos os genitores

Considere o seguinte conjunto de dados e genealogia, associados à avaliação de uma espécie frutífera.

Descendentes	Bloco	Pai	Mãe	Frutos (kg)
4	1	1	-	22,7
5	2	3	2	14,6
6	2	1	2	19,7
7	1	4	5	17,7
8	1	3	6	25,3

10.7.2.1 Modelo individual – MI

Para a obtenção das predições BLUP dos valores genéticos dos indivíduos deve-se resolver o seguinte sistema de equações de modelo misto - EMM:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ em que } \alpha = \sigma_e^2 / \sigma_a^2 = (1 - h^2) / h^2.$$

As matrizes de incidência para os efeitos fixos (efeitos de bloco) e aleatórios (valores genéticos dos indivíduos) são:

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

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

Note-se por Z, que, embora os indivíduos 1, 2 e 3 não possuam observação, pode-se avaliá-los geneticamente com base em informações de seus descendentes.

Sendo $y = [22,7 \ 14,6 \ 19,7 \ 17,7 \ 25,3]$, obtém-se os seguintes produtos matriciais:

$$X'X = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

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

$$Z'y = [0 \ 0 \ 0 \ 22,7 \ 14,6 \ 19,7 \ 17,7 \ 25,3]'$$

$$X'y = \begin{bmatrix} 65,7 \\ 34,3 \end{bmatrix}$$

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

Considerando as relações de parentesco genético aditivo entre os indivíduos em avaliação, tem-se que a inversa (A^{-1}) da matriz de parentesco é:

$$A^{-1} = \begin{bmatrix} 1.833 & 0.500 & 0.000 & -0.667 & 0.000 & -1.000 & 0.000 & 0.000 \\ 0.500 & 2.000 & 0.500 & 0.000 & -1.000 & -1.000 & 0.000 & 0.000 \\ 0.000 & 0.500 & 2.000 & 0.000 & -1.000 & 0.500 & 0.000 & -1.000 \\ -0.667 & 0.000 & 0.000 & 1.833 & 0.500 & 0.000 & -1.000 & 0.000 \\ 0.000 & -1.000 & -1.000 & 0.500 & 2.500 & 0.000 & -1.000 & 0.000 \\ -1.000 & -1.000 & 0.500 & 0.000 & 0.000 & 2.500 & 0.000 & -1.000 \\ 0.000 & 0.000 & 0.000 & -1.000 & -1.000 & 0.000 & 2.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & -1.000 & 0.000 & 2.000 \end{bmatrix}$$

Considerando uma herdabilidade de 0,33, $\alpha = (1-h^2)/h^2 = 2$, de forma que $A^{-1}\alpha$ é obtida multiplicando-se cada elemento de A^{-1} por 2. Somando $A^{-1}\alpha$ a $Z'Z$, as EMM são:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 3.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 4.000 & 0.000 & -2.000 & 1.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & -1.333 & 0.000 & 0.000 & 4.667 & 1.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & -2.000 & -2.000 & 1.000 & 6.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & -2.000 & -2.000 & 1.000 & 0.000 & 0.000 & 6.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & -2.000 & -2.000 & 0.000 & 5.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & -2.000 & 0.000 & 0.000 & -2.000 & 0.000 & 5.000 \end{bmatrix}^{-1} \begin{bmatrix} 65,7 \\ 34,3 \\ 0 \\ 0 \\ 0 \\ 22,7 \\ 14,6 \\ 19,7 \\ 17,7 \\ 25,3 \end{bmatrix}$$

Resolvendo-se, tem-se:

Efeitos	Soluções
Bloco	
$\hat{\beta}_1$	22,0274
$\hat{\beta}_2$	17,1677
Indivíduos	
\hat{a}_1	0,4979
\hat{a}_2	-0,0908
\hat{a}_3	-0,2057
\hat{a}_4	-0,0535
\hat{a}_5	-0,9386
\hat{a}_6	0,9032
\hat{a}_7	-1,2623
\hat{a}_8	0,9335

10.7.2.2. Modelo individual reduzido – MIR

No modelo individual reduzido, proposto inicialmente por Quaas & Pollak (1980), apenas equações para os genitores são construídas.

Para os indivíduos não genitores, a_d e e_2 podem ser combinados em um único resíduo, $e^* = a_d + e_2$.

Em notação matricial e tomando-se $e_2 = e$, os modelos são (Resende & Rosa-Perez, 1999):

Modelo individual

$$y = Xb + Za + e$$

Modelo individual reduzido para os genitores

$$y_g = X_g b + Z a_g + e$$

Modelo individual reduzido para os não genitores

$$y_n = X_n b + Z_1 a_g + e^*, \text{ em que:}$$

X_n : matriz de incidência para efeitos fixos, referentes aos não genitores.

Z_1 : matriz de incidência, cujos únicos elementos não zero equivalem a $\frac{1}{2}$, identificando os genitores dos indivíduos (o valor $\frac{1}{2}$ advém da consideração de que $a - a_d = \frac{1}{2} (a_p + a_m)$).

a_g : vetor de valores genéticos dos genitores.

e^* : vetor de erros aleatórios : $e^* = a_d + e$.

Assim, o modelo individual reduzido - MIR - pode ser escrito como:

$$\begin{bmatrix} y_g \\ y_n \end{bmatrix} = \begin{bmatrix} X_g \\ X_n \end{bmatrix} b + \begin{bmatrix} Z \\ Z_1 \end{bmatrix} a_g + \begin{bmatrix} e \\ e^* \end{bmatrix}$$

Denominando-se:

$$X = \begin{bmatrix} X_g \\ X_n \end{bmatrix}, \quad W = \begin{bmatrix} Z \\ Z_1 \end{bmatrix} \quad R = \begin{bmatrix} R_g \\ R_n \end{bmatrix} = \begin{bmatrix} I\sigma_e^2 & 0 \\ 0 & I\sigma_{e^*}^2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I + D^* \alpha^{-1} \end{bmatrix} \sigma_e^2$$

as estruturas de variância são:

$$Var(y) = WA_g W' \sigma_a^2 + R$$

$$Var(a_g) = A_g \sigma_a^2$$

em que:

A_g : é a matriz de parentesco para genitores e D^* é uma matriz diagonal com elementos iguais a d_j , em que $d_j = \frac{1}{2}$, $\frac{3}{4}$ ou 1, se ambos, um ou nenhum dos genitores são conhecidos, respectivamente.

$$\sigma_{e^*}^2 = \sigma_e^2 + d_j \sigma_a^2 = (1+d_j \alpha^{-1}) \sigma_e^2, \text{ ignorando a endogamia.}$$

$$\alpha = \sigma_e^2 / \sigma_a^2$$

Assim, as equações de modelo misto são:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}W \\ W'R^{-1}X & W'R^{-1}W + A_g^{-1}(1/\sigma_a^2) \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ W'R^{-1}y \end{bmatrix}$$

Considerando o conjunto de dados, os indivíduos 1, 2, 3, 4, 5 e 6 são genitores enquanto os indivíduos 7 e 8 não são genitores. Considerando os parâmetros genéticos $h^2 = 0,33$, $\alpha = 2$, $\sigma_a^2 = 2,75$ e $\sigma_e^2 = 5,5$, os elementos da diagonal de R são iguais a $\sigma_e^2 = 5,5$ para os indivíduos genitores (4, 5 e 6) e a $(\sigma_e^2 + d_i \sigma_a^2) = 5,5 + (1/2) 2,75 = 6,875$ para os indivíduos não genitores (7 e 8). Assim,

$$R = \begin{bmatrix} 5,5 & 0 & 0 & 0 \\ 5,5 & 0 & 0 & 0 \\ & 5,5 & 0 & 0 \\ Simétrica & & 6,875 & 0 \\ & & & 6,875 \end{bmatrix}$$

$$R^{-1} = \begin{bmatrix} 0,1818 & 0 & 0 & 0 & 0 \\ & 0,1818 & 0 & 0 & 0 \\ & & 0,1818 & 0 & 0 \\ Sim & & & 0,1455 & 0 \\ & & & & 0,1455 \end{bmatrix}$$

X é exatamente como no modelo individual, ou seja:

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

1 2 3 4 5 6 (indivíduos)

$$W = \begin{bmatrix} Z \\ Z_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0,5 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0 & 0,5 \end{bmatrix}$$

4 5 6 7 8

$$y = [22,7 \ 14,6 \ 19,7 \ 17,7 \ 25,3]'$$

Efetuando-se os produtos matriciais, têm-se os elementos necessários nas EMM:

$$X'R^{-1}X = \begin{bmatrix} 0,4727 & 0 \\ 0 & 0,3636 \end{bmatrix}$$

$$X'R^{-1}W = \begin{bmatrix} 0 & 0 & 0,0727 & 0,2545 & 0,0727 & 0,0727 \\ 0 & 0 & 0 & 0 & 0,1818 & 0,1818 \end{bmatrix}$$

$$W'R^{-1}X = [X'R^{-1}W]'$$

$$W'R^{-1}W = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,0364 & 0 & 0 & 0,0364 \\ 0 & 0 & 0 & 0,2182 & 0,0364 & 0 \\ 0 & 0 & 0 & 0,0364 & 0,2182 & 0 \\ 0 & 0 & 0,0364 & 0 & & 0,2182 \end{bmatrix}$$

$$X'R^{-1}Y = \begin{bmatrix} 10,3818 \\ 6,2364 \end{bmatrix}$$

$$W'R^{-1}y = \begin{bmatrix} 0,0000 \\ 0,0000 \\ 1,8400 \\ 5,4145 \\ 3,9418 \\ 5,4218 \end{bmatrix}$$

A inversa da matriz de parentesco somente para os indivíduos genitores, isto é, indivíduos 1 a 6 é:

$$A^{-1} = \begin{bmatrix} 1,833 & 0,500 & 0,000 & -0,667 & 0,000 & -1,000 \\ 0,500 & 2,000 & 0,500 & 0,000 & -1,000 & -1,000 \\ 0,000 & 0,500 & 1,500 & 0,000 & -1,000 & 0,000 \\ -0,667 & 0,000 & 0,000 & 1,333 & 0,000 & 0,000 \\ 0,000 & -1,000 & -1,000 & 0,000 & 2,000 & 0,000 \\ -1,000 & -1,000 & 0,000 & 0,000 & 0,000 & 2,000 \end{bmatrix}$$

Somando-se $A^{-1}(1/\sigma_a^2)\mathbf{a}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}$ e reunindo-se os demais elementos, obtém-se as EMM:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 0,4727 & 0 & 0 & 0 & 0,0727 & 0,2545 & 0,0727 & 0,0727 \\ 0 & 0,3636 & 0 & 0 & 0 & 0 & 0,1818 & 0,1818 \\ 0 & 0 & 0,6665 & 0,1818 & 0 & -0,2425 & 0 & -0,3636 \\ 0 & 0 & 0,1818 & 0,7273 & 0,1818 & 0 & -0,3636 & -0,3636 \\ 0,0727 & 0 & 0 & 0,1818 & 0,5818 & 0 & -0,3636 & 0,0364 \\ 0,2545 & 0 & -0,2425 & 0 & 0 & 0,7029 & 0,0364 & 0 \\ 0,0727 & 0,1818 & 0 & -0,3636 & -0,3636 & 0,0364 & 0,9455 & 0 \\ 0,0727 & 0,1818 & -0,3636 & -0,3636 & 0,0364 & 0 & 0 & 0,9455 \end{bmatrix}^{-1} \begin{bmatrix} 10,3818 \\ 6,2364 \\ 0 \\ 0 \\ 1,8400 \\ 5,4145 \\ 3,9418 \\ 5,4218 \end{bmatrix}$$

Resolvendo, obtém-se:

Efeitos	Soluções
Blocos	
\hat{b}_1	22,0274
\hat{b}_2	17,1677
Genitores	
\hat{a}_1	0,4979
\hat{a}_2	-0,0908
\hat{a}_3	-0,2057
\hat{a}_4	-0,0535
\hat{a}_5	-0,9386
\hat{a}_6	0,9032

Estas soluções são exatamente iguais àquelas obtidas pelo modelo individual. Entretanto, o esforço computacional foi menor, uma vez que o número de equações para os efeitos aleatórios reduziu-se de 8 para 6.

As soluções para os indivíduos não genitores são resolvidas com base na equação:

$$\hat{a} = 0,5 (\hat{a}_p + \hat{a}_m) + a_d$$

$$= 0,5 (a_p + a_m) + h_d^2 (y - X\hat{b} - 0,5 \hat{a}_p - 0,5 \hat{a}_m), \text{ em que}$$

$h_d^2 = (1/2 \sigma_a^2)/(1/2 \sigma_a^2 + \sigma_e^2)$: herdabilidade dentro de progênies de irmãos germanos, que equivale a 0,20 no presente exemplo.

O cálculo de \hat{a}_7 e \hat{a}_8 usando esta expressão é dado por:

$$\hat{a}_7 = 0,5 (\hat{a}_4 + \hat{a}_5) + h_d^2 (17,7 - \hat{b}_1 - 0,5 \hat{a}_4 - 0,5 \hat{a}_5) = -1,2623$$

$$\hat{a}_8 = 0,5 (\hat{a}_3 + \hat{a}_6) + h_d^2 (25,3 - \hat{b}_1 - 0,5 \hat{a}_3 - 0,5 \hat{a}_6) = 0,9335$$

Estas soluções são exatamente iguais àquelas obtidas pelo modelo individual.

De maneira genérica, $h_d^2 = (dh^2)/(dh^2 + (1-h^2))$, equivalendo a $h_d^2 = (1/2 h^2)/(1/2 h^2 + (1-h^2))$ para progênies de irmãos germanos e $h_d^2 = (3/4 h^2)/(3/4 h^2 + (1-h^2))$ para progênies de meios-irmãos.

10.7.3. Modelos Computacionais quando se conhece apenas um genitor

Considere o seguinte conjunto de dados e genealogia, associados à avaliação de uma espécie florestal, através de famílias de meios-irmãos, em que cada 2 árvores constituem uma parcela.

Indivíduo	Bloco	Família (Mãe)	Árvore	Diâmetro (cm)
4	1	1	1	9,87
5	1	1	2	14,48
6	1	2	1	8,91
7	1	2	2	14,64
8	1	3	1	9,55
9	1	3	2	7,96
10	2	1	1	16,07
11	2	1	2	14,01
12	2	2	1	7,96
13	2	2	2	21,17
14	2	3	1	10,19
15	2	3	2	9,23

10.7.3.1 Modelo individual – MI

Nesta situação, o modelo linear misto adequado para a descrição dos dados equivale a:

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

y , b , a , c e e : vetores de dados, de efeitos fixos (médias de blocos), de efeitos aditivos (aleatório), de efeitos de parcelas (efeitos aleatórios de ambiente comum das parcelas) e de erros aleatórios, respectivamente.

X , Z e W : são matrizes de incidência conhecidas, formadas por valores 0 e 1, às quais associam as incógnitas b , a e c ao vetor de dados y , respectivamente.

A metodologia de modelos mistos permite estimar b pelo procedimento de quadrados mínimos generalizados e predizer a e c pelo procedimento BLUP. Para a obtenção destas soluções, basta resolver o seguinte sistema de equações lineares, o qual é denominado equações de modelo misto (MME):

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

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

A e I = matrizes de parentesco genético aditivo e matriz identidade de ordem apropriada aos dados, respectivamente.

Empregando-se os dados apresentados no tópico anterior, têm-se as seguintes matrizes de incidência:

$$X_{(12x2)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad Z_{(12x15)} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad W_{(12x6)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

A matriz I no caso é de ordem 6 e a matriz A com a inclusão das 3 matrizes equivale a:

$$A = \begin{bmatrix} 1.0000 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 & 0 & 0 & 0 & 0.5000 & 0.5000 & 0 & 0 & 0 & 0 & 0 & 5.000 & 5.000 \\ 0.5000 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0 \\ 0.5000 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 \\ 0.5000 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 & 0 & 0 \\ 0.5000 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 & 0 & 0 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 1.0000 & 0.2500 & 0 \\ 0 & 0 & 0.5000 & 0 & 0 & 0 & 0 & 0.2500 & 0.2500 & 0 & 0 & 0 & 0 & 0.2500 & 1.0000 & 0 \end{bmatrix}$$

Considerando os parâmetros $h^2 = 0,1635$ e $c^2 = 0,0779$, Obtém-se $\lambda_1 = 4,639$ e $\lambda_2 = 9,738$. Resolvendo as MME, obtém-se o vetor de soluções:

$$\begin{bmatrix} 10,9017 \\ 13,1050 \\ 0,4370 \\ 0,3178 \\ -0,7548 \\ 0,0224 \\ 0,6641 \\ -0,1554 \\ 0,6422 \\ -0,4760 \\ \hat{b} \\ \hat{a} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} -0,6973 \\ 0,5649 \\ 0,2782 \\ -0,6065 \\ 1,2321 \\ -0,6675 \\ -0,8011 \\ 0,1585 \\ 0,1073 \\ -0,2658 \\ 0,2579 \\ 0,1955 \\ -0,4534 \end{bmatrix}$$

No vetor de soluções, os dois primeiros valores referem-se às estimativas BLUE (melhor estimativa linear não viciada) das médias dos blocos 1 e 2, respectivamente, os três valores seguintes são relativos aos efeitos genéticos aditivos preditos para as 3 matrizes, os 12 valores seguintes dizem respeito aos efeitos genéticos aditivos preditos das progênies (descendentes) e os 6 últimos valores, aos efeitos preditos para as parcelas.

10.7.3.2. Modelo individual reduzido – MIR

No modelo reduzido, o mesmo sistema de equações se aplica, porém o vetor de efeitos genéticos aditivos contempla apenas os genitores. No caso em que não se têm observações (dados) referentes a genitoras, as EMM se reduzem a:

$$\begin{bmatrix} \hat{b} \\ \hat{a}_g \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'W \\ Z_1'X & Z_1'Z_1 + A_g^{-1}\lambda_1 & Z_1'W \\ W'X & W'Z & W'W + I\lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2 + (3/4)\sigma_a^2}{\sigma_a^2} = \frac{4 - h^2 - 4c^2}{4h^2} = 5,389755 \text{ no presente exemplo;}$$

$$\lambda_2 = \frac{\sigma_e^2 + (3/4)\sigma_a^2}{\sigma_c^2} = \frac{4 - h^2 - 4c^2}{4c^2} = 11,31226 \text{ no presente exemplo.}$$

As matrizes X , W e y são exatamente como especificado no tópico 10.7.3.1. A matriz A , no caso de testes de progénie de genitoras não aparentadas, é uma matriz identidade (neste exemplo, de ordem 3).

A matriz Z_1 é composta por valores 0 e 0,5, equivalendo a:

$$Z_{l(12x3)} = \begin{bmatrix} 0,5 & 0 & 0 \\ 0,5 & 0 & 0 \\ 0 & 0,5 & 0 \\ 0 & 0,5 & 0 \\ 0 & 0 & 0,5 \\ 0 & 0 & 0,5 \\ 0,5 & 0 & 0 \\ 0,5 & 0 & 0 \\ 0 & 0,5 & 0 \\ 0 & 0,5 & 0 \\ 0 & 0 & 0,5 \\ 0 & 0 & 0,5 \end{bmatrix},$$

O vetor de soluções deste exemplo é equivalente a:

$$\begin{bmatrix} \hat{b} \\ \hat{a}_g \\ \hat{c} \end{bmatrix} = \begin{bmatrix} 10,9017 \\ 13,1050 \\ 0,4370 \\ 0,3178 \\ -0,7548 \\ 0,1585 \\ 0,1073 \\ -0,2658 \\ 0,2579 \\ 0,1955 \\ -0,4534 \end{bmatrix}$$

Verifica-se que esse modelo fornece como resultados os valores genéticos integrais das genitoras bem como os efeitos de bloco e parcelas, exatamente como no modelo MI.

Para cômputo dos valores genéticos dos indivíduos, deve-se empregar a expressão $\hat{a} = Z_1 \hat{a}_g + h_d^2(y - X\hat{b} - Z_1 \hat{a}_g - W\hat{c})$, a qual fornece os seguintes resultados, que são idênticos aos obtidos pelo MI:

$$\hat{a} = \begin{bmatrix} 0,0224 \\ 0,6641 \\ -0,1554 \\ 0,6422 \\ -0,4760 \\ -0,6973 \\ 0,5649 \\ 0,2782 \\ -0,6065 \\ 1,2321 \\ -0,6675 \\ -0,8011 \end{bmatrix}$$

A herdabilidade dentro de progênies de meios-irmãos para um modelo com parcelas de várias plantas é dada por $h_d^2 = (3/4)h^2/(3/4)h^2 + (1 - h^2 - c^2)$ e equivale a 0,13915 no presente caso. O valor individual corrigido para os efeitos de bloco, parcela e progénie é ponderado pela própria h_d^2 .

As soluções propiciadas pela expressão $\hat{a} = Z_1 \hat{a}_g + h_d^2(y - X\hat{b} - Z_1 \hat{a}_g - W\hat{c})$ são equivalentes às obtidas pelo método do índice multiefeitos (Resende & Higa, 1994), o qual foi implementado no software Selegen (Resende et al., 1994). De fato, quando os dados são balanceados, os métodos BLUP e índice multiefeitos são idênticos, ou seja, nesta situação o método do índice multiefeitos é BLUP.

Como exemplo, o resultado obtido para o primeiro indivíduo é dado por $\hat{a} = 0,5 * 0,4370 + 0,13915(9,87 - 10,9017 - 0,5 * 0,4370 - 0,1585) = 0,0224$.

Esta formulação é também benéfica à estimação REML de componentes de variância. Neste caso, os seguintes estimadores REML podem ser usados:

$$\hat{\sigma}^2 = [y'y - \hat{b}'X'y - \hat{a}'_g Z_1'y - \hat{c}'W'y]/[N - r(x)] = \hat{\sigma}_e^2 + (3/4)\hat{\sigma}_a^2$$

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

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

tr: operador traço matricial; r(x): posto da matriz X; q : número de genitoras; s : número de parcelas; N : número total de dados; C²² e C³³ provêm de:

$$C = \begin{bmatrix} C^{11} & C^{12} & C^{13} \\ C^{21} & C^{22} & C^{23} \\ C^{31} & C^{32} & C^{33} \end{bmatrix} = \text{inversa generalizada da matriz dos coeficientes das EMM.}$$

Para o caso de testes de progênies de irmãos germanos com várias plantas por parcela, a mesma formulação do MIR pode ser aplicada, bastando considerar que:

$$\lambda_1 = \frac{\sigma_e^2 + (1/2)\sigma_a^2}{\sigma_a^2} = \frac{2 - h^2 - 2c^2}{2h^2};$$

$$\lambda_2 = \frac{\sigma_e^2 + (1/2)\sigma_a^2}{\sigma_c^2} = \frac{2 - h^2 - 2c^2}{2c^2};$$

$$h_d^2 = (1/2 h^2) / (1/2 h^2 + (1 - h^2 - c^2)).$$

10.7.3.3. Modelo gamético – MG

No modelo gamético, o mesmo sistema de equações se aplica, porém o vetor de efeitos genéticos aditivos contempla apenas metade do valor genético dos genitores. As EMM são:

$$\begin{bmatrix} \hat{b} \\ (\hat{a}/2) \\ \hat{c} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A_g^{-1}\lambda_1 & Z'W \\ W'X & W'Z & W'W + I\lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}, \text{ em que:}$$

$$\lambda_1 = \frac{\sigma_e^2 + (3/4)\sigma_a^2}{(1/4)\sigma_a^2} = \frac{4 - h^2 - 4c^2}{h^2} = 21,5590 \text{ no presente exemplo;}$$

$$\lambda_2 = \frac{\sigma_e^2 + (3/4)\sigma_a^2}{\sigma_c^2} = \frac{4 - h^2 - 4c^2}{4c^2} = 11,31226 \text{ no presente exemplo.}$$

As matrizes X, W e y são exatamente como especificado no tópico 7.10.3.1. A matriz A, no caso de testes de progénie de genitoras não aparentadas, é uma matriz identidade (neste exemplo, de ordem 3).

A matriz Z_1 é composta por valores 0 e 1, equivalendo a:

$$Z_{l(12x3)} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

O vetor de soluções deste exemplo é equivalente a:

$$\begin{bmatrix} \hat{b} \\ (\hat{a}/2) \\ \hat{c} \end{bmatrix} = \begin{bmatrix} 10,9017 \\ 13,1050 \\ 0,2185 \\ 0,1589 \\ -0,3774 \\ 0,1585 \\ 0,1073 \\ -0,2658 \\ 0,2579 \\ 0,1955 \\ -0,4534 \end{bmatrix}$$

Verifica-se que este modelo fornece como resultados a metade dos valores genéticos das genitoras bem como os efeitos de bloco e parcelas. A variância de $(a/2)$ contempla $(1/4)$ da variância genética aditiva, que equivale à variância genética entre progênies de meios-irmãos.

Para cômputo dos valores genéticos dos indivíduos, deve-se empregar a expressão $\hat{a} = Z(\hat{a}/2) + h_d^2(y - X\hat{b} - Z(\hat{a}/2) - W\hat{c})$, a qual fornece resultados idênticos aos obtidos pelo MI e MIR.

10.7.4. Modelo de clones repetidos – MCR

Considere a avaliação de clones não aparentados no delineamento de blocos ao acaso com várias plantas por parcela e uma medição por indivíduo.

Neste caso, o modelo linear misto adequado à descrição dos dados equiva-
le a:

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

y , b , g , c e e : vetores de dados, de efeitos fixos (médias de blocos), de efeitos genotípicos (aleatórios), de efeitos de parcelas (aleatórios) e de erros aleatórios, respectivamente.

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

As equações de modelo misto para a estimativa dos efeitos fixos e predi-
ção dos efeitos aleatórios equivalem a:

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

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

Usando o mesmo conjunto de dados apresentados no tópico 10.7.3 e os parâmetros $h_a^2 = 0,2453$ e $c^2 = 0,0779$, tem-se $\lambda_1 = 2,75907$, $\lambda_2 = 8,68806$ e obtêm-se os efeitos genotípicos preditos equivalentes a 0,8678; 0,6311 e -1,4989, para os clones 1, 2 e 3, respectivamente. No caso, a matriz Z é exata-
mente igual à do modelo gamético e Var (g) contempla a variância genotípica total.

O uso dos modelos MIR, MG e MCR propicia redução de cerca de 85% na dimensão dos sistemas de equações lineares, ou seja, reduz em 85% o esforço computacional necessário. Por exemplo, em um teste com 100 progênies, 60 indivíduos por progénie e 10 blocos, o sistema de equações reduz-se de 7.110

para 1.110 equações. Esta redução pode ser aplicada tanto nos modelos univariados quanto multivariados (análise conjunta de locais considerando a interação genótipo x ambiente). Além disso, em casos mais simples, como progêneres de polinização aberta e testes clonais envolvendo genitores não aparentados, nem é necessária a construção da matriz de parentesco (A). Extensões para o caso de progêneres pertencentes a várias populações e para o delineamento em látice são triviais. Modelos reduzidos foram utilizados na implementação do software Selegen–REML.

Uma redução adicional no sistema de equações pode ser conseguida por meio da absorção (eliminação de Gauss) do efeito de parcela, fazendo-se:

$$K = I_N - W (W' W + I \lambda_2)^{-1} W' : \text{matriz de absorção dos efeitos de parcela.}$$

$$\begin{bmatrix} \hat{b} \\ \hat{a}_g \end{bmatrix} = \begin{bmatrix} X' KX & X' KZ \\ Z' KX & Z' KZ + A_g^{-1} \lambda_1 \end{bmatrix}^{-1} \begin{bmatrix} X' Ky \\ Z' Ky \end{bmatrix}.$$

A matriz $(W' W + I \lambda_2)$ é diagonal e possui inversa trivial. Dessa forma, no exemplo mencionado, após a absorção, ocorre uma redução de 7.110 para 100 equações, ou seja, 98,6%. Posteriormente, os efeitos de parcela podem ser obtidos por $\hat{c} = (W' W + I \lambda_2)^{-1} (W' y - W' Z \hat{a}_g - W' X \hat{b})$

Outra estratégia é absorver também os efeitos fixos e resolver diretamente as equações, sem o uso de métodos iterativos. Neste caso, absorvendo b em a_g , obtém-se:

$$X' M X \hat{b} = X' M y, \text{ em que:}$$

$$M = K - K Z (Z' K Z + A_g \lambda_1)^{-1} Z' K : \text{matriz de absorção dos efeitos fixos.}$$

Sendo $A_g = I_q$, a matriz $(Z' K Z + A_g \lambda_1)$ é diagonal e possui inversa trivial. Assim, as soluções podem ser obtidas diretamente por:

$$\begin{aligned} \hat{b} &= (X' M X)^{-1} X' M y \\ \hat{a}_g &= (Z' K Z + I \lambda_1)^{-1} (Z' Ky - Z' K X \hat{b}) \\ \hat{c} &= (W' W + I \lambda_2)^{-1} (W' y - W' Z \hat{a}_g - W' X \hat{b}) \end{aligned}$$

As absorções mencionadas facilitam também a obtenção de estimativas REML dos componentes de variância. Neste caso, os traços de C^{22} e C^{33} podem ser obtidos usando matrizes que surgem durante os passos da absorção, portanto, sem a necessidade de inverter a matriz dos coeficientes C. Assim, têm-se:

$$\text{tr} (C^{22}) = \text{tr} (H_a) + \text{tr} (H_b L_{xa} L'_{xa})$$

$\text{tr} (C^{33}) = \text{tr} (H_c) + \text{tr} (H_a L_{ac} L'_{ac}) + \text{tr} (H_b T)$, em que:

$$H_b = (X'M X)^{-1}$$

$$H_a = (Z'K Z + I \lambda_1)^{-1}$$

$$H_c = (W'W + I \lambda_2)^{-1}$$

$$L_{xa} = X'KZ H_a$$

$$L_{xc} = X'W H_c$$

$$L_{ac} = Z'W H_c$$

$$T = [L_{xc} \quad -L_{xa} \quad L_{ac}] \begin{bmatrix} L'_{xc} \\ -L'_{ac} \quad L'_{xa} \end{bmatrix}$$

Desse modo, para cômputo indireto dos traços, 3 matrizes simétricas são necessárias:

- (a) $L_{ac} L'_{ac}$, de ordem igual ao número de níveis de a_g ;
- (b) $L_{xa} L'_{xa}$, de ordem igual ao número total de níveis dos efeitos fixos;
- (c) T , de ordem igual ao número total de níveis dos efeitos fixos.

Capítulo 11

**Regressão Aleatória e Funções
de Covariâncias na Análise
de Dados Longitudinais**

11.1. Introdução

Nos modelos gerais de regressão (simples ou múltipla), as covariáveis ou variáveis independentes na realidade não são variáveis, mas sim constantes, ou seja, em hipotética amostragem repetida os valores dessas covariáveis permanecem constantes. Nestes modelos, os coeficientes de regressão são fixos, isto é, os mesmos coeficientes de regressão estão associados a todos os indivíduos da população. Existem modelos nos quais a suposição adequada referente aos coeficientes de regressão associados às covariáveis é que cada membro possui distribuição aleatória. Estes modelos são denominados modelos de regressão aleatória.

A teoria associada à regressão aleatória foi apresentada, inicialmente, por Henderson Júnior (1982). Com base nesta teoria existem, em um dado experimento, diferentes coeficientes de regressão associados aos diferentes indivíduos da população, de forma que se os indivíduos são efeitos aleatórios, os coeficientes de regressão associados a estes indivíduos devem também ser tratados como efeitos aleatórios (Henderson, 1984).

Como no caso do melhoramento genético de espécies perenes, os indivíduos são tratados como efeitos aleatórios por meio de seus valores genéticos. Também, de maneira geral, avaliações repetidas são realizadas em cada indivíduo no decorrer do tempo ou idade, de forma que variáveis respostas podem ser explicadas mediante uma função linear da idade. Neste contexto, os modelos de regressão aleatória são capazes de explicar melhor a variabilidade associada a medidas repetidas do que os modelos lineares simples (que incluem o efeito de ambiente permanente e o parâmetro repetibilidade ou tratam as medidas repetidas como caracteres diferentes por meio de um modelo multivariado) empregados em avaliação genética. Os modelos de regressão aleatória permitem a obtenção de diferentes curvas de valores genéticos associadas aos diferentes indivíduos e consideram as mudanças nas variâncias genéticas e residuais através do tempo. Também, desde que existam confiáveis estimativas de componentes de variância, estes modelos permitem a predição de valores genéticos de um indivíduo nas diferentes idades, baseando-se em avaliações de apenas uma idade.

No melhoramento genético de espécies perenes, a regressão aleatória no contexto dos modelos lineares mistos apresenta grande utilidade, pois permite a predição de valores genéticos de indivíduos avaliados em diferentes idades (e com diferentes números de idades avaliadas) e a projeção destes valores genéticos para uma idade comum, para efeito de ordenamento e seleção. Apesar de a aplicação dos modelos de regressão aleatória ser comum em outras áreas (Longford, 1993), a aplicação desses modelos ao melhoramento genético iniciou-se, recentemente, com os trabalhos de Schaeffer & Dekkers (1994), Jamrozick & Schaeffer (1997) e Jamrozick et al. (1997), visando à avaliação genética em gado bovino leiteiro.

A estimação de covariâncias entre os coeficientes de regressão aleatória produz estimativas de funções de covariâncias (Kirkpatrick et al., 1990), que dizem respeito a uma descrição contínua da estrutura de covariâncias do caráter para a amplitude de idades abrangida pelos dados. Em outras palavras, uma função de covariância descreve a covariância entre medições tomadas em certas idades como função destas idades. Meyer (1998a) demonstrou a equivalência entre os modelos de regressão aleatória e de funções de covariância.

Os modelos de regressão aleatória e as funções de covariância podem expressar, de maneira mais realista, os fenômenos associados aos dados longitudinais (por exemplo, curvas de crescimento através do tempo ou medições repetidas). Meyer (1998a) relata que estes modelos deverão suplantar os modelos menos realistas até então usados. Os modelos de regressão aleatória permitem considerar que o caráter em questão pode estar mudando continuamente e gradualmente no tempo (Kirkpatrick et al. 1994) e, portanto, podem ser mais realísticos que os modelos de repetibilidade e multivariado.

A seguir são apresentados aspectos metodológicos associados aos modelos de regressão aleatória e às possibilidades de aplicação ao melhoramento genético de espécies perenes, tomando por base os trabalhos de Resende (1999b) e Resende et al. (2001a).

11.2. Regressão Aleatória no Contexto dos Modelos Lineares Mistos

Para caracteres avaliados em vários estágios de desenvolvimento dos indivíduos, o seguinte modelo linear pode ser adotado (para o caso univariado):

$$Y_{ijk} = P_i + (b_0 + b_1 A^*)_j + (a + \lambda A^*)_j + e_{ijk}, \text{ onde:}$$

Y_{ijk} : é a k-ésima medição do indivíduo j associado ao efeito fixo i ($i = 1, 2, 3, \dots, b$; $j = 1, 2, 3, \dots, n$; $k = 1, 2, 3, \dots, m$).

P_i : representa os efeitos fixos do modelo.

b_0 e b_1 : coeficientes de regressão fixa da medição na idade.

A^* : idade em meses ou anos.

a e λ : coeficientes de regressão aleatória relacionando a medição do indivíduo j a um intercepto e idade, respectivamente. Em conjunto são os efeitos aleatórios de valor genético aditivo. a representa diferenças genéticas entre indivíduos não explicadas pela idade e λ representa diferenças genéticas entre indivíduos em função da idade.

e_{ijk} : erro aleatório.

Por este modelo, a curva de valores genéticos para determinado indivíduo pode ser vista mediante dois grupos de regressão: regressão fixa para todos os indivíduos pertencendo a um mesmo nível de efeito fixo, a qual descreve a forma geral para um determinado indivíduo; regressões aleatórias descrevendo os desvios da regressão fixa, gerando diferentes curvas para os diferentes indivíduos.

Em notação matricial, o modelo pode ser escrito:

$y = X\beta + Z_0 a + Z_1 \lambda + e$, onde:

y , β , e : vetores de dados, de efeitos fixos e de erros aleatórios, respectivamente.

a , λ : vetores referentes aos coeficientes de regressão aleatória, que relacionam as medições aos interceptos e idades, respectivamente.

X : matriz de incidência para os efeitos fixos.

Z_0 : matriz de incidência para a , contendo 0 e 1's.

Z_1 : matriz associando λ a Y , contendo zero e valores de idade.

Nas matrizes Z_0 e Z_1 as primeiras colunas são nulas devido ao fato de que referem-se aos indivíduos que não possuem observação.

São as seguintes as médias e a estrutura de variâncias associadas ao modelo:

$E(y) = X\beta$ e $E(Y_{ijk}) = P_i + (b_0 + b_1 A^*)$, em que $(b_0 + b_1 A^*)$ refere-se à média geral na idade A^* .

$$\text{Var} \begin{pmatrix} a \\ \lambda \\ e \end{pmatrix} = \begin{bmatrix} Ag_{aa} & Ag_{a\lambda} & 0 \\ Ag_{a\lambda} & Ag_{\lambda\lambda} & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}, \text{ onde:}$$

A : matriz de parentesco genético aditivo.

g_{ij} : covariâncias entre os coeficientes de regressão aleatória, ou seja, são funções de covariância que descrevem continuamente a estrutura de covariâncias para o caráter, no intervalo de idades abrangido pelos dados.

A matriz de covariância (G) entre os efeitos genéticos aleatórios, desconsiderando as relações de parentesco equivale a:

$$G = \begin{pmatrix} g_{aa} & g_{a\lambda} \\ g_{a\lambda} & g_{\lambda\lambda} \end{pmatrix}$$

11.3. Estimação dos Efeitos Fixos e Predição dos Efeitos Aleatórios

A estimação e predição dos efeitos segundo o modelo linear definido é realizada pelas equações de modelo misto:

$$\begin{bmatrix} X'X & X'Z_0 & X'Z_1 \\ Z_0'X & Z_0'Z_0 + A^{-1}\alpha_{00} & Z_0'Z_1 + A^{-1}\alpha_{01} \\ Z_1'X & Z_1'Z_0 + A^{-1}\alpha_{01} & Z_1'Z_1 + A^{-1}\alpha_{11} \end{bmatrix} \begin{pmatrix} \beta \\ a \\ \lambda \end{pmatrix} = \begin{pmatrix} X'y \\ Z_0'y \\ Z_1'y \end{pmatrix}, \text{ onde:}$$

$$\alpha_{ij} = g^{ij} \sigma_e^2$$

g^{ij} : provém de G^{-1}

σ_e^2 : variância residual média

β : vetor de efeitos fixos

Assim:

$$G^{-1} \sigma_e^2 = \begin{pmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{pmatrix}$$

11.4. Estimação dos Componentes de Variância e Covariância

As variâncias genética e fenotípica são dependentes da idade, ou seja, podem aumentar ou diminuir com a idade. Tem-se que:

$$\hat{\sigma}_{a(i)}^2 = k_i' \hat{G} k_i = g_{aa} + 2A^* g_{a\lambda} + A^{*2} g_{\lambda\lambda} : \text{variância genética na idade } i.$$

$$\hat{\sigma}_{y(i)}^2 = \hat{\sigma}_{a(i)}^2 + \hat{\sigma}_e^2 : \text{variância fenotípica na idade } i.$$

$$\hat{\sigma}_{a(ij)}^2 = k_i' \hat{G} k_j = g_{aa} + (A_i^* + A_j^*) g_{a\lambda} + A_i^* A_j^* g_{\lambda\lambda} : \text{covariância genética entre as idades } i \text{ e } j.$$

$$r_{a(ij)} = \frac{\hat{\sigma}_{a(ij)}^2}{\hat{\sigma}_{a(i)}^2 \hat{\sigma}_{a(j)}^2} : \text{é a correlação genética entre as idades } i \text{ e } j.$$

$k_i' = [1 \text{ idade } i]$: transposto do vetor que contempla o valor 1 e o valor da idade i .

Para a utilização das equações de modelo misto e estimação das variâncias, covariâncias e correlações genéticas são necessárias as estimativas de G e σ_e^2 .

Meyer (1997), Shah et al. (1997) e Meyer & Hill (1997) estudaram aspectos da estimação de parâmetros genéticos para dados longitudinais por máxima verossimilhança restrita – REML. Com base no algoritmo EM, são os seguintes os estimadores para os componentes de variância:

$$\hat{\sigma}_e^2 = (y' y - \hat{\beta}' X' y - \hat{a}' Z_0' y - \hat{\lambda}' Z_1' y) / (N - r(x))$$

$$\hat{g}_{ij} = (\hat{\theta}_i A^{-1} \hat{\theta}_j + \text{tr } A^{-1} C_{ij} \sigma_e^2) / q$$

onde:

θ_i : a_i ou λ_i .

N : número total de observações.

$r(x)$: posto de X ou número de colunas linearmente independentes de X .

q: número de elementos aleatórios (número de indivíduos ou valores genéticos a serem preditos).

tr: operação traço matricial ou soma dos elementos da diagonal da matriz.

C_{ij} : submatriz que provém da inversa generalizada (C) da matriz dos coeficientes das equações de modelo misto:

$$C = \begin{bmatrix} C_{xx} & C_{xz_0} & C_{xz_I} \\ C_{z_0x} & C_{z_0z_0} & C_{z_0z_I} \\ C_{z_Ix} & C_{z_Iz_0} & C_{z_Iz_I} \end{bmatrix}$$

Com base nos valores iniciais para G e σ_e^2 obtém-se β , \mathbf{a} e λ a partir das equações de modelo misto, os quais são utilizados nos estimadores $\hat{\sigma}_e^2$ e \hat{g}_{ij} , cujas estimativas são retornadas às equações de modelo misto. Procede-se, desta maneira, sucessivamente, até a convergência.

Valores iniciais de g_{aa} , $g_{a\lambda}$ e $g_{\lambda\lambda}$ podem ser obtidos a partir de estimativas de $\sigma_{a(i)}^2$ para três idades, conseguidas pelo método de máxima verossimilhança restrita.

11.5. Predição dos Valores Genéticos em Cada Idade

Os valores genéticos para o indivíduo j podem ser preditos para várias idades (i) por meio da regressão linear aleatória:

$$\hat{VG}_{ij} = \hat{\mathbf{a}}_j + \hat{\lambda}_j \text{ (idade } i\text{)}$$

As previsões \hat{VG}_{ij} permitem a ordenação dos candidatos à seleção, nas diferentes idades, possibilitando a seleção de indivíduos de acordo com a idade desejada.

11.6. Exemplo Numérico com Dados Simulados

O presente exemplo considera a avaliação de 7 indivíduos (alguns avaliados em mais de uma idade) pertencentes a 2 populações, para o caráter densidade da madeira em diferentes idades, conforme especificado na Tabela 11.1.

Tabela 11.1. Genealogia e dados referentes à avaliação de 7 indivíduos, para o caráter densidade da madeira, em diferentes idades.

Indiví-duo	Pai	Mãe	População	Idade (meses)	Densidade da madeira (kg/0,1m ³)
4	1	-	1	4	20
4	1	-	1	15	35
5	3	2	2	8	27
5	3	2	2	10	30
6	1	2	2	6	21
6	1	2	2	12	32
7	4	5	2	13	28
7	4	5	2	15	32

A seguinte estrutura de variâncias e covariâncias foi considerada:

$$\sigma_e^2 = 5,50$$

$$G = \begin{bmatrix} 2,7500 & -0,0581 \\ -0,0581 & 0,0550 \end{bmatrix}$$

$$G^{-1}\sigma_e^2 = \begin{bmatrix} 2,0457 & 2,1610 \\ 2,1610 & 102,2828 \end{bmatrix}$$

Com base nos componentes de variância apresentados têm-se, na Tabela 11.2, as estimativas de variância genética aditiva (σ_a^2), variância fenotípica (σ_y^2) e herdabilidade (h^2) em função da idade.

Tabela 11.2. Estimativas da variância genética aditiva (σ_a^2), variância fenotípica (σ_y^2) e herdabilidade (h^2) em função da idade (meses).

Idade	σ_a^2	σ_y^2	(h^2)
4	3,1652	8,6652	0,3653
5	3,5440	9,0440	0,3919
6	4,0328	9,5328	0,4230
7	4,6316	10,1316	0,4571
8	5,3404	10,8404	0,4926
9	6,1592	11,6592	0,5283
10	7,0880	12,5880	0,5631
11	8,1268	13,6268	0,5964
12	9,2756	14,7756	0,6278
13	10,5344	16,0344	0,6570
14	11,9032	17,4032	0,6840
15	13,3820	18,8820	0,7087

Verifica-se nesse exemplo que a variação genética e a herdabilidade aumentaram com o passar da idade, refletindo a mudança contínua do caráter em função do tempo, para a faixa de idades considerada.

Na Tabela 11.3 são apresentadas as covariâncias ($\sigma_{a_{ij}}$) e correlações genéticas ($r_{a_{ij}}$) entre alguns pares de idades.

Tabela 11.3. Estimativas das covariâncias ($\sigma_{a_{ij}}$) e correlações genéticas ($r_{a_{ij}}$) entre alguns pares de idades.

Par de Idades	Covariância genética ($\sigma_{a_{ij}}$)	Correlação genética ($r_{a_{ij}}$)
4-5	3,3271	0,9934
4-15	4,9461	0,7600
14-15	12,6151	0,9995

Verificam-se altas correlações genéticas entre todas as combinações de idades, das quais as mais altas são as estimativas relativas às medidas adjacentes, conforme esperado.

As matrizes de incidência são:

$$\mathbf{x} = \begin{bmatrix} 1 & 0 & 4 \\ 1 & 0 & 15 \\ 0 & 1 & 8 \\ 0 & 1 & 10 \\ 0 & 1 & 6 \\ 0 & 1 & 12 \\ 0 & 1 & 13 \\ 0 & 1 & 15 \end{bmatrix}, \text{ considerando um modelo de médias.}$$

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

$$Z_t = \begin{bmatrix} 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 15 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 15 \end{bmatrix}$$

A matriz A^{-1} equivale a:

$$A^{-1} = \begin{bmatrix} 1,8333 & 0,5000 & 0 & -0,6667 & 0 & -1,0000 & 0 \\ 0,5000 & 2,0000 & 0,5000 & 0 & -1,0000 & -1,0000 & 0 \\ 0 & 0,5000 & 1,5000 & 0 & -1,0000 & 0 & 0 \\ -0,6667 & 0 & 0 & 1,8333 & 0,5000 & 0 & -1,0000 \\ 0 & -1,0000 & -1,0000 & 0,5000 & 2,5000 & 0 & -1,0000 \\ -1,0000 & -1,0000 & 0 & 0 & 0 & 2,0000 & 0 \\ 0 & 0 & 0 & -1,0000 & -1,0000 & 0 & 2,0000 \end{bmatrix}$$

Resolvendo as equações de modelo misto, obtém-se as soluções, conforme a Tabela 11.4.

Tabela 11.4. Soluções das equações de modelo misto.

Efeitos	Soluções
Fixos	
$\hat{P}_1 + \hat{b}_0$	15,3375
$\hat{P}_2 + \hat{b}_0$	13,5377
\hat{b}_1	1,4179
Aleatórios a	
\hat{a}_1	-0,2663
\hat{a}_2	0,2006
\hat{a}_3	0,2663
\hat{a}_4	-0,4341
\hat{a}_5	0,4998
\hat{a}_6	-0,0986
\hat{a}_7	-0,3443

Continua...

Tabela 11.4. Continuação.

Efeitos	Soluções
Aleatórios λ	
$\hat{\lambda}_1$	-0,0252
$\hat{\lambda}_2$	0,0529
$\hat{\lambda}_3$	0,0252
$\hat{\lambda}_4$	-0,0919
$\hat{\lambda}_5$	0,0642
$\hat{\lambda}_6$	0,0416
$\hat{\lambda}_7$	-0,1425

Com base nestas soluções, têm-se as seguintes estimativas:

Média da população 1 aos 4 meses:

$$P_1 + b_0 + b_1 A^* = 15,3375 + 1,4179 \times 4 = 21,0091$$

Média da população 1 aos 15 meses:

$$P_1 + b_0 + b_1 A^* = 15,3375 + 1,4179 \times 15 = 36,6060$$

Média da população 2 aos 4 meses:

$$P_2 + b_0 + b_1 A^* = 13,5377 + 1,4179 \times 4 = 19,2093$$

Média da população 2 aos 15 meses:

$$P_2 + b_0 + b_1 A^* = 13,5377 + 1,4179 \times 15 = 34,8062$$

Os efeitos genéticos aditivos dos indivíduos, em diferentes idades, são apresentados na Tabela 11.5 e são dados por (VG do indivíduo 1 com idade de 4 meses):

$$VG_{14} = \hat{a}_1 + \hat{\lambda}_1 \times 4 = -0,2663 + (-0,0252) \times 4 = -0,3671 \text{ e, assim, sucessivamente.}$$

Tabela 11.5. Efeitos genéticos preditos dos indivíduos em diferentes idades.

Indivíduo	Efeito genético			
	Idade = 4 meses	Idade = 8 meses	Idade = 12 meses	Idade = 15 meses
1	-0,3671	-0,4679	-0,5687	-0,6443
2	0,4122	0,6238	0,8353	0,9940
3	0,3670	0,4676	0,5683	0,6437
4	-0,8018	-1,1695	-1,5372	-1,8130
5	0,7565	1,0133	1,2701	1,4626
6	0,0678	0,2342	0,4006	0,5254
7	-0,9141	-1,4840	-2,0538	-2,4811

A metodologia permite, então, a obtenção de curvas de valores genéticos para cada indivíduo em função da idade, bem como a comparação dos vários indivíduos a uma idade comum, mesmo quando alguns indivíduos são avaliados em uma ou poucas idades, no intervalo de idades considerado.

11.7. Aplicação a Dados Experimentais

Foram utilizados dados do caráter circunferência à altura do peito, obtidos a partir de um teste de 45 progênies de *Eucalyptus urophylla*, conduzido no município de Aracruz-ES, pela empresa Aracruz Celulose S.A. O delineamento experimental empregado foi o de blocos ao acaso com 9 repetições e 6 plantas por parcela, e o referido caráter foi avaliado em 7 idades (1 a 7 anos). Foram utilizados 14.278 dados, os quais foram analisados pelo programa DXMRR (Meyer, 1998b), disponível no aplicativo DFREML. Adotou-se um modelo que inclui os efeitos aditivos (1 função de covariância com 3 componentes de variância: g_{aa} ; $g_{a\lambda}$ e $g_{\lambda\lambda}$), os efeitos de ambiente permanente (1 função de covariância com 3 componentes de variância) e os efeitos residuais (7 componentes de variância do resíduo, ou seja, um para cada idade). Assim, o modelo de regressão aleatória esteve associado a 13 componentes de variância.

Foram comparadas 3 classes de modelos de análise: modelos univariados para cada idade separadamente; modelos de repetibilidade com as várias idades; modelos de regressão aleatória. Os resultados dos modelos univariados são apresentados na Tabela 11.6.

Tabela 11.6. Estimativas* de parâmetros genéticos em análises univariadas para o caráter circunferência à altura do peito em *Eucalyptus urophylla* em várias idades, bem como o número (N) de indivíduos avaliados em cada idade, a sobrevivência (S) e as repetibilidades (r_{ij}) para o caráter sobrevivência relativo as idades i e j .

Idade	Herdabilidade	Variância aditiva	N	S	$i-j$	r_{ij}
1	$0,150 \pm 0,04$	2,6665	2326	0,9572	1-2	0,97
2	$0,186 \pm 0,05$	9,8157	2255	0,9280	2-3	0,95
3	$0,237 \pm 0,06$	18,3412	2144	0,8823	3-4	0,94
4	$0,357 \pm 0,09$	36,7589	2017	0,8300	4-5	0,96
5	$0,369 \pm 0,09$	48,8212	1936	0,7967	5-6	0,95
6	$0,378 \pm 0,09$	54,9871	1839	0,7568	6-7	0,96
7	$0,390 \pm 0,09$	78,1604	1767	0,7272	1-7	0,76
					2-7	0,78
					3-7	0,82
					4-7	0,88
					5-7	0,91
					6-7	0,96

* Obtidas pelo método da máxima verossimilhança restrita – REML.

Verifica-se, pela Tabela 11.6, aumento progressivo da herdabilidade e redução progressiva da sobrevivência com o aumento da idade.

Na Tabela 11.7 são apresentados os resultados associados ao modelo de regressão aleatória.

Tabela 11.7. Estimativas* de parâmetros genéticos pelo modelo de regressão aleatória para o caráter circunferência à altura do peito em *Eucalyptus urophylla* nas idades de 1 a 7 anos.

Idade	Herdabilidade	Variância aditiva	Par de idades	Correlação genética	Repetibilidade
1	$0,153 \pm 0,02$	2,7475	1-2	0,862	0,909
2	$0,176 \pm 0,03$	8,9071	1-3	0,756	0,868
3	$0,211 \pm 0,03$	21,3117	1-4	0,697	0,834
4	$0,232 \pm 0,04$	39,9610	1-5	0,661	0,811
5	$0,245 \pm 0,05$	64,8552	1-6	0,637	0,790
6	$0,251 \pm 0,05$	95,9942	1-7	0,620	0,767
7	$0,251 \pm 0,05$	133,3780	2-3	0,983	0,947
			2-4	0,964	0,940
			2-5	0,950	0,933
			2-6	0,939	0,921
			2-7	0,932	0,904
			3-4	0,996	0,981
			3-5	0,991	0,981
			3-6	0,986	0,973
			3-7	0,982	0,959
			4-5	0,999	0,994
			4-6	0,997	0,989
			4-7	0,995	0,976
			5-6	0,999	0,996
			5-7	0,999	0,984
			6-7	1,0	0,982

* Baseada em 14.278 dados de 2.326 indivíduos.

A Tabela 11.7 mostra que a herdabilidade aumenta ligeiramente com o passar da idade, estabilizando-se entre os 3 e 5 anos de idade. As estimativas obtidas para as idades de 1, 2 e 3 anos são próximas às obtidas pelas análises univariadas (Tabela 11.6), mas as obtidas para as demais idades não o são.

A provável explicação para essas diferenças refere-se ao efeito da reduzida sobrevivência a partir (a menor repetibilidade para a sobrevivência em idades adjacentes, ocorreu para a combinação 3-4 anos, conforme a Tabela 11.6) do 4º ano e seus reflexos nas metodologias de análise. Por exemplo, para o 4º ano, a técnica da regressão aleatória baseou-se em 14.278 dados de 2.326 indivíduos, ao passo que a análise univariada, em apenas 2.017 dados de 2.017 indivíduos. Assim, a análise univariada estima parâmetros para uma população com reduzida variância fenotípica devido à morte dos indivíduos menos vigorosos. A técnica da regressão aleatória utiliza a correlação entre idades para aproveitar a informação dos indivíduos que morreram em idades prévias, estimando, assim, os parâmetros para toda a população inicial e não para a população já selecionada naturalmente para adaptação. Esses efeitos da sobrevivência nas estimativas de parâmetros genéticos foram também apresentados por Matheson & Raymond (1984) em *Pinus radiata* e também por meio de simulação. Tais autores reportaram que as estimativas de herdabilidade para as populações completas foram de 0,12 e 0,24 e passaram a 0,21 e 0,33 após a eliminação das piores plantas, para 2 populações de *Pinus radiata*, respectivamente.

Dessa forma, os resultados obtidos por modelos univariados para cada idade não são precisos. À medida que os anos passam, os parâmetros passam a ser representativos de uma população selecionada para adaptação e são, portanto, superestimados conduzindo a inferências incorretas sobre os valores genéticos e ganhos genéticos. Entretanto, se avaliações de idades precoces (2 ou 3 anos) são utilizadas para a seleção, os erros pelo uso dos modelos univariados, provavelmente, são desprezíveis (Tabelas 11.6 e 11.7).

Um caráter avaliado repetidas vezes sobre um mesmo indivíduo pode ser considerado como um mesmo caráter ao longo das idades se as correlações genéticas entre medidas repetidas deste caráter forem iguais a 1 (Henderson, 1984). As estimativas de correlações genéticas apresentadas na Tabela 11.7 mostram que a circunferência à altura do peito pode ser considerada um mesmo caráter a partir dos três anos (correlações $> 0,982$), idade esta que coincide com o início da estabilização da herdabilidade do caráter. Pode-se inferir, com base nas correlações, que o crescimento ao longo das idades equivale, praticamente, a 3 caracteres: crescimento a 1 ano; crescimento aos 2 anos; crescimento a partir de 3 anos. Os resultados referentes às estimativas de repetibilidade (correlações fenotípicas entre idades) apresentados na Tabela 11.7 e incrementos médios anuais (Tabela 11.8) corroboram esta inferência.

As estimativas de repetibilidade (Tabela 11.7) indicam que um modelo de repetibilidade pode ser adequado para dados de crescimento a partir dos 3 anos (pois a amplitude das estimativas de repetibilidade é reduzida para esta faixa de idades). Entretanto, para adoção do modelo de repetibilidade é mais adequado considerar a variável incremento anual. Os resultados dos modelos de repetibilidade para o incremento anual da circunferência são apresentados na Tabela 11.8.

Tabela 11.8. Estimativas de parâmetros genéticos empregando modelos de repetibilidade para o caráter incremento anual da circunferência à altura do peito em *Eucalyptus urophylla*, bem como as médias de idade e incrementos médios anuais.

Idades	Herdabilidade	Repetibili-dade	Variância aditiva	Média (cm)	Incremento médio anual (cm)
1 a 7 anos	0,180 \pm 0,04	0,67	2,7466	-	-
3 a 7 anos	0,210 \pm 0,05	0,69	1,9249	-	-
1 e 2 anos	0,126 \pm 0,04	0,57	2,2071	-	-
1	-	-	-	14,21	14,21
2	-	-	-	25,97	11,76
3	-	-	-	31,08	5,11
4	-	-	-	35,73	4,65
5	-	-	-	38,94	3,21
6	-	-	-	41,37	2,43
7	-	-	-	45,99	4,62

Observa-se que tratando o incremento anual na circunferência como um único caráter, este apresentou herdabilidade de 0,18 e repetibilidade de 0,67. Entretanto, tal variável não deve ser tratada como um único caráter, já que herdabilidades e repetibilidades diferentes foram obtidas nas faixas de idade de 1 a 2 anos e de 3 a 7 anos. Os valores de incremento médio anual também são mais próximos a partir do 3º incremento e, nos dois primeiros anos a taxa de crescimento é muito superior (Tabela 11.8). Com base nestes resultados e naqueles da Tabela 11.7, há evidência de que o modelo de repetibilidade pode ser aplicado adequadamente a partir do 3º incremento.

Em função da pequena importância da avaliação de 1 ano e do fato de que dados nesta idade raramente são tomados na prática, pode-se recomendar a predição dos valores genéticos em duas etapas: aos 2 anos para circunferência, empregando-se a análise univariada; para incrementos anuais a partir do 2º ano pelo modelo de repetibilidade. Dessa forma, os valores genéticos preditos para a idade de 7 anos são dados $\hat{a}_2 + 5\hat{a}_m$, em que \hat{a}_2 referem-se aos valores genéticos preditos para circunferência aos 2 anos e \hat{a}_m , aos valores genéticos preditos para o incremento anual.

Em resumo, na análise quantitativa de medidas repetidas, 4 classes de modelos de análise podem ser empregadas: (i) modelos univariados para cada idade; (ii) modelos de repetibilidade; (iii) modelos de regressão aleatória e (iv) modelos multivariados. Destes, os mais completos são os modelos multivariados, que utilizam toda a informação, simultaneamente, e tratam cada idade como um caráter, considerando suas diferentes herdabilidades e correlações genéticas. Entretanto, tais modelos são superparametrizados e, na prática, proibitivos quando muitas idades são consideradas, tal como no presente trabalho. Isto ocorre devido ao fato de que, neste caso, problemas de estimação surgem principalmente quando o verdadeiro valor do parâmetro encontra-se próximo

(como no presente trabalho em que as correlações foram muito próximas a 1) ao limite do espaço paramétrico (Meyer & Hill, 1997), acarretando problemas computacionais e produzindo estimativas imprecisas. Assim, devem ser preferidos modelos com menor número de parâmetros e que descrevam, adequadamente, os dados observados.

Os modelos de regressão aleatória podem ser considerados como modelos multivariados simplificados ou reduzidos, que permitem a obtenção dos mesmos parâmetros de interesse obtidos pelos modelos multivariados, mas com menor parametrização. No presente trabalho, um modelo multivariado requereria 56 parâmetros (28 covariâncias aditivas, 21 covariâncias de ambiente permanente e 7 variâncias residuais), ao passo que a regressão aleatória empregou 13 parâmetros e, portanto, mostrou-se uma técnica muito informativa e adequada.

Com relação aos outros dois modelos, neste trabalho, o modelo de repetibilidade mostrou-se adequado para os dados a partir do 3º incremento anual e o modelo univariado mostrou-se adequado para dados até o 3º ano. Empregados nestas situações estes modelos, embora simplificados, são úteis e eficientes.

Os modelos mais completos (multivariado e de regressão aleatória) devem ser utilizados de acordo com o número de idades disponíveis para análise: com dados de 2 idades (por exemplo 2 e 3 anos), o modelo multivariado pode ser utilizado de maneira eficiente; com dados de 4 ou mais idades, o modelo de regressão aleatória pode ser utilizado adequadamente.

Em geral, pode-se concluir que:

- A técnica de regressão aleatória permite: descrever continuamente a estrutura de covariâncias de caracteres de crescimento ao longo do tempo; a estimação simultânea de parâmetros genéticos nas várias idades, considerando a mudança genética do caráter com o tempo; a predição de valores genéticos de indivíduos avaliados em diferentes idades e a projeção destes valores genéticos para uma idade comum para efeitos de ordenamento e seleção; a obtenção de curvas de valores genéticos para cada indivíduo em função da idade.
- O modelo de regressão aleatória na análise genética pode conduzir a melhores resultados dos que os tradicionais modelos univariados e de repetibilidade, devendo, assim, ser o preferido.
- O modelo de regressão aleatória permite a obtenção de estimativas dos mesmos parâmetros que podem ser obtidos por um modelo multivariado, porém com uma menor parametrização, fato que é vantajoso computacionalmente.
- Finalmente, é importante relatar que, embora o modelo de regressão aleatória tenha sido muito recomendado, tal modelo não é, necessariamente, adequado para todos os conjuntos de dados. Outras opções existem para a modelagem de dados longitudinais, conforme apresentado no tópico seguinte.

11.8. Síntese dos Métodos de Análise de Dados Longitudinais no Melhoramento

Os caracteres de interesse no melhoramento de plantas perenes se expressam mais de uma vez em cada indivíduo, gerando dados longitudinais ou medidas repetidas. Tais caracteres são denominados infinitamente dimensionais. Dados desses caracteres apresentam estrutura correlacionada através do tempo, safras ou medições. Algumas alternativas de análise desses caracteres são:

(a) Análise univariada considerando cada idade ou safra em separado (estrutura de covariância não correlacionada).

Neste caso, as avaliações em diferentes estágios são consideradas como sendo caracteres diferentes e não correlacionados. Assim, os dados de cada idade são analisados separadamente. Nesta situação, a estrutura geral (válida para efeitos genéticos e ambientais) de covariância entre quatro diferentes idades ou safras é dada por:

$$\sum = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A estrutura de covariância genética é dada por:

$$\sum_g = \begin{bmatrix} \sigma_{g_1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{g_2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{g_3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{g_4}^2 \end{bmatrix}, \text{em que a } \sigma_{g_i}^2 \text{ é a variância genética na idade } i.$$

Este modelo não é realista e deve ser evitado. Em algumas plantas perenes, uma só avaliação, às vezes, é considerada. Por exemplo, em espécies florestais como o eucalipto, a seleção precoce é enfatizada procurando-se uma idade mínima que torne a seleção eficiente. Em geral, a maioria dos programas conduzidos com esta espécie no Brasil utiliza eficientemente a seleção baseada em uma só avaliação aos três anos.

Em outras plantas perenes como as fruteiras, muitas vezes, é gerado apenas um dado por indivíduo, dado este referente à soma ou média de várias safras. A seleção baseada apenas nesse dado totalizado ou médio raramente será completamente eficiente, a menos que todos os indivíduos tenham apre-

sentado produção em todas as safras. Por outro lado, a seleção baseada em médias por grupo de plantas e em várias safras, como a seleção de genitores baseada na progênie, a seleção de clones em testes clonais e a seleção de híbridos ou famílias híbridas, tende a ser menos prejudicada pelo uso de um único dado médio.

(b) Análise pelo modelo de repetibilidade (estrutura de covariância completamente correlacionada).

Nesta situação, considera-se que o caráter é o mesmo através das várias safras. Isto implica assumir que a correlação genética entre qualquer par de safras equivale a 1, que a variância fenotípica é a mesma para todas as safras e que a correlação ambiental permanente é a mesma para todos os pares de safras. Essas premissas são atendidas aproximadamente para algumas fruteiras, mas não para todas. Todo o conjunto de dados é analisado simultaneamente e a estrutura de covariância genética entre quatro diferentes safras é dada por:

$$\sum_g = \sigma_g^2 \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \text{ em que } \sigma_g^2 \text{ é a variância genética.}$$

A estrutura geral de covariância entre quatro diferentes safras é dada por:

$$\sum = \begin{bmatrix} k & a & a & a \\ a & k & a & a \\ a & a & k & a \\ a & a & a & k \end{bmatrix},$$

em que k representa os elementos diagonais (variâncias) constantes e a representa os valores de correlação entre pares de safras ou idade.

Em espécies florestais, este modelo de análise é raramente aplicável. Isto porque, com o aumento dos caracteres de crescimento com a idade, as suposições de igual variância e correlação não são realistas. Neste caso, a padronização prévia dos dados remove apenas parcialmente a heterogeneidade de variâncias.

(c) Modelo multivariado (Matriz de Covariância não Estruturada).

Este é o modelo mais completo, utiliza toda a informação simultaneamente e trata idades ou safras como sendo caracteres diferentes e correlacionados, considerando suas diferentes herdabilidades e correlações genéticas. A estrutura geral de covariância entre quatro diferentes idades é dada por:

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

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

(d) Modelo de Regressão Aleatória (Estrutura de covariância definida por funções de covariância).

O modelo de regressão aleatória pode ser considerado como um modelo multivariado reduzido e simplificado, o qual permite a obtenção dos mesmos parâmetros de interesse (herdabilidade em cada idade e correlação genética entre todos os pares de idade), que podem ser obtidos pelo modelo multivariado, porém com uma menor parametrização e com menor esforço computacional. Tal abordagem define diretamente funções de covariância contínuas e permite incluir na análise indivíduos com idades heterogêneas. A estrutura geral de covariância entre quatro diferentes idades é dada por:

$$\sum = \begin{bmatrix} z_{01} & z_{11} & z_{21} & z_{31} \\ z_{02} & z_{12} & z_{22} & z_{32} \\ z_{03} & z_{13} & z_{23} & z_{33} \\ z_{04} & z_{14} & z_{24} & z_{34} \end{bmatrix} \begin{bmatrix} 1 & a & b & c \\ a & 1 & d & e \\ b & d & 1 & f \\ c & e & f & 1 \end{bmatrix} \begin{bmatrix} z_{01} & z_{02} & z_{03} & z_{04} \\ z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \\ z_{31} & z_{32} & z_{33} & z_{34} \end{bmatrix}$$

$$= Q_i \Lambda_0 Q_i^T, \text{ em que:}$$

z_{ij} = i-ésimo vetor polinomial ortogonal analisado na idade variável j .

Λ_0 = matriz de covariância dos regressores aleatórios.

Q_i = matriz com $q+1$ colunas contendo $z_0, z_1, z_2, \dots, z_q$, respectivamente, em que q é a ordem do polinômio e z_1 é o i-ésimo vetor polinomial ortogonal.

Um modelo de regressão aleatória pode ser ajustado somente com um intercepto e inclinação, com um termo quadrático adicional, com um termo cúbico adicional e assim sucessivamente de acordo com a ordem do polinômio de Legendre ajustado.

(e) Modelo Auto-regressivo com Variâncias Heterogêneas (Estrutura Auto-regressiva).

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

$$\sum = \begin{bmatrix} 1 & a^{|t_2-t_1|} & a^{|t_3-t_1|} & a^{|t_4-t_1|} \\ a^{|t_2-t_1|} & 1 & a^{|t_3-t_2|} & a^{|t_4-t_2|} \\ a^{|t_3-t_1|} & a^{|t_3-t_2|} & 1 & a^{|t_4-t_3|} \\ a^{|t_4-t_1|} & a^{|t_4-t_2|} & a^{|t_4-t_3|} & 1 \end{bmatrix}$$

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

(f) Modelo com correlações específicas para cada intervalo de idade
(Estrutura de Covariância Toeplitz).

Este modelo considera correlações iguais para iguais intervalos entre idades e variâncias heterogêneas entre idades. É denominado modelo de correlação bandada e a matriz de covariância entre idades tem estrutura Toeplitz, que é dada por:

$$\sum = \begin{bmatrix} 1 & a & b & c \\ a & 1 & a & b \\ b & a & 1 & a \\ c & b & a & 1 \end{bmatrix}$$

Uma matriz Toeplitz apresenta várias diagonais (primárias, secundárias, terciárias etc.). A matriz de covariância genética é dada por $G = S \Sigma S$, em que Σ , no caso, é uma matriz Toeplitz.

As estruturas auto-regressivas (AR) e Toeplitz impõem fortes restrições às covariâncias, gerando também modelos mais parcimoniosos, assim como a regressão aleatória.

(g) Ajuste de uma curva *spline* cúbica no intervalo de idades considerado.

Uma *spline* cúbica é uma curva alisada ou suavizada sobre um intervalo $[a,b]$ formado pela ligação de segmentos polinomiais cúbicos a “pontos nó” tais que a curva inteira e suas primeiras e segundas diferenças são contínuas sobre o intervalo. Refere-se a um processo de interpolação que propicia uma curva

contínua e suave a partir dos pontos observados, sendo diferenciável e integrável para um domínio de interesse.

A função *spline* cúbica usa um processo polinomial de terceiro grau para a interpolação de valores entre cada par de pontos observados. Neste caso, uma polinomial diferente é usada para cada intervalo e cada uma é construída de forma a sempre passar pelos dados originais e apresentar uma derivada contínua nas junções entre cada intervalo. Uma curva definida por uma polinomial cúbica pode passar exatamente por quatro pontos, sendo que, no caso de uma longa seqüência de pontos, torna-se necessário usar uma sucessão de intervalos polinomiais. Para garantir que não ocorram mudanças súbitas nas inclinações ou curvaturas entre os intervalos sucessivos, tal função polinomial não é ajustada para quatro pontos e sim para dois.

A função *spline* cúbica foi desenvolvida inicialmente por Shoenberg (1946) e aperfeiçoada por Walsh et al. (1962). Verbyla et al. (1999) e White et al. (1999) sugeriram o uso de funções mais flexíveis, tais quais polinomiais de elevada ordem ou *splines* cúbicas para modelar dados longitudinais quando não existe conhecimento prévio sobre o modelo biológico adjacente ao caráter.

Exemplos e detalhes da obtenção de *splines* cúbicas são apresentados por Davis (1986) e Green & Silverman (1994).

Uma comparação entre estas várias abordagens, exceto (b) e (g), na estimação de parâmetros genéticos em *Pinus radiata* foi realizada por Apiolaza et al. (2000). Tais autores relataram que os modelos de regressão aleatória não são necessariamente adequados para qualquer conjunto de dados, tendo verificado que os modelos simples, como o AR, com variâncias específicas para cada idade, parecem ser flexíveis o suficiente para serem usados com eficiência em muitas situações no melhoramento florestal. Pelo modelo AR, o valor genético de um indivíduo i na época j é dado por $a_{ij} = \rho a_{ij-1} + \alpha_j$, em que a_{ij-1} é o valor genético no tempo prévio $j-1$, ρ é o coeficiente de autocorrelação serial e α_j é o efeito genético adicional na nova medição j .

Capítulo 12

**Inferência Estatística Bayesiana
e Simulação Estocástica
na Avaliação Genética**

12.1. Sinopse

A predição de valores genéticos e a estimativa de componentes de variância são atividades essenciais no melhoramento de plantas perenes, sendo que o procedimento padrão de estimação/predição é o REML/BLUP – máxima verossimilhança restrita/melhor predição linear não viciada. Uma deficiência do procedimento REML/BLUP para a estimação/predição de componentes de variância/valores genéticos refere-se ao fato de que o método REML propicia intervalos de confiança apenas aproximados para os parâmetros genéticos, por meio do uso de aproximações e suposições de normalidade assintótica. Isto porque a distribuição e a variância dos estimadores não são conhecidas e, assim, questões referentes à efetividade da seleção a ser praticada não podem ser respondidas com rigor.

A análise bayesiana baseia-se no conhecimento da distribuição a posteriori dos parâmetros genéticos e possibilita a construção de intervalos de confiança (melhor definidos como intervalos de probabilidade ou intervalos de confiança bayesianos) exatos para as estimativas dos parâmetros genéticos. Assim, esta análise propicia uma descrição mais completa sobre a confiabilidade dos parâmetros genéticos do que o método REML (Gianola & Fernando, 1986; Resende, 1997; 1999b).

A inferência estatística bayesiana baseia-se na distribuição condicional do parâmetro (θ) dado o vetor de dados (y), ou seja, na distribuição a posteriori do parâmetro dadas as observações fenotípicas, a qual equivale a:

$$f(\theta|y) = \frac{f(y|\theta) f(\theta)}{\int_R f(y|\theta) f(\theta) d\theta}, \text{ em que:}$$

$f(y|\theta)$: função densidade de probabilidade da distribuição condicional de um vetor de observações (y) dado θ (denominada função de verossimilhança ou modelo para os dados).

$f(\theta)$: função densidade de probabilidade da distribuição a priori, que é também a densidade marginal de θ . Esta função denota o grau de conhecimento acumulado sobre θ , antes da observação de y .

$\int_R f(y|\theta) f(\theta) d\theta$: distribuição marginal ou preditiva de y com respeito à θ , em que R é a amplitude da distribuição de θ .

Verifica-se, então, que a distribuição a posteriori é proporcional à verossimilhança x priori, ou seja, a função de verossimilhança conecta a priori à posteriori usando para isto os dados do experimento (observações). Dessa forma, a distribuição a posteriori contempla o grau de conhecimento prévio sobre o parâmetro (θ) e também as informações adicionais propiciadas pelo experimento (y).

Para a situação em que tanto a distribuição a priori quanto a distribuição das informações são normais, ou seja, $\theta \sim N(u, r^2)$ e $(Y|\theta) \sim N(\theta, \sigma^2)$ com σ^2 conhecido, a distribuição a posteriori de θ é também normal, ou seja $(\theta|Y = y) \sim N(u_1, r_1^2)$ onde:

$$\mu_1 = \frac{r^{-2}\mu + \sigma^{-2}y}{r^{-2} + \sigma^{-2}}; \quad r_1^{-2} = r^{-2} + \sigma^{-2}$$

Usando esta distribuição a posteriori, pode-se demonstrar que os fundamentos básicos da predição de valores genéticos são essencialmente de natureza bayesiana, conforme apresentado inicialmente por Robertson (1955).

Os efeitos genéticos aditivos são definidos como desvios e uma população de efeitos genéticos apresenta zero como média e σ_a^2 como variância. A melhor predição do efeito genético de um indivíduo sem nenhuma informação, tomado aleatoriamente da população é a média populacional u , a qual pode ser tomada como o estimador a priori, cuja variância é σ_a^2 . Tomando uma informação y do indivíduo, um segundo estimador do efeito genético é o desvio $(y-u)$ fenotípico em relação à média populacional, o qual possui variância $\sigma_e^2 = \sigma_y^2 - \sigma_a^2$, em que σ_y^2 é a variância de y . Estes dois estimadores independentes podem ser combinados linearmente da melhor maneira possível, tomando as recíprocas das respectivas variâncias como pesos. Sob o enfoque bayesiano, a esperança da distribuição a posteriori corresponde à média ponderada pela precisão, das médias da priori e da verossimilhança. Assim, tem-se:

$$\begin{aligned} \hat{a} &= \left(\frac{\text{estimador a priori}}{\sigma_a^2} + \frac{\text{estimador informação observada}}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = \\ &= \left(\frac{0}{\sigma_a^2} + \frac{Y-u}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = h^2 (y-u), \end{aligned}$$

como no enfoque da inferência estatística clássica, em que h^2 é a herdabilidade. Este é também (considerando y como a média dos indivíduos selecionados) o estimador para o ganho genético para a seleção massal.

No contexto bayesiano, as equações de modelo misto (denominadas equações de modelo misto de Robertson) são dadas por:

$$\begin{bmatrix} X'R^{-1}X + S^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} E(\beta|y) \\ E(a|y) \end{bmatrix} = \begin{bmatrix} X'R^{-1}y + S^{-1}r_1 \\ Z'R^{-1}y + G^{-1}0 \end{bmatrix}$$

em que $r_1 = E(\beta)$ e $0 = E(a)$.

Tomando a informação a priori sobre os efeitos fixos (β) como não informativa (expressa como $S \rightarrow \infty$ e então $S^{-1} \rightarrow 0$), tem-se que esta equação resultante equivale às equações do modelo misto – EMM:

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

$R = I\sigma_e^2$; $G = A\sigma_a^2$ e $S = \text{Var}(\beta)$; A : matriz de correlação genética aditiva entre os indivíduos em avaliação.

O problema básico da implementação da análise bayesiana refere-se à integração numérica (no espaço do parâmetro) da função densidade de probabilidade a posteriori. Tal integração, por métodos analíticos, é impossível na prática. Assim, tem sido utilizado o procedimento de simulação estocástica denominado amostragem de Gibbs (GS) para a viabilização da estimativa bayesiana (Sorensen, 1996).

Além das distribuições (normais) adotadas para os efeitos aleatórios (a) no modelo linear clássico e para a verossimilhança do vetor de observações (y), a abordagem bayesiana requer atribuições para as distribuições a priori dos efeitos fixos e componentes de variância. De maneira geral, são atribuídas distribuições a priori não informativas ou uniformes para os efeitos fixos e componentes de variância, como forma de caracterizar um conhecimento a priori vago sobre os referidos efeitos e componentes (Gianola & Fernando, 1986).

Utilizando-se distribuições a priori não informativas para os efeitos fixos e componentes de variância, as modas das distribuições marginais a posteriori dos parâmetros genéticos correspondem às estimativas obtidas por REML e geram estimativas de componentes de variância pelo método VEIL (ou de verossimilhança integrada) apresentado por Gianola & Foulley (1990). A grande vantagem da análise bayesiana, neste caso, refere-se à obtenção dos desvios padrões e intervalos de confiança exatos para os parâmetros genéticos bem como a obtenção de estimativas mais precisas. Empregando as abordagens REML e GS, Resende (1999b) obteve estimativas de $0,32 \pm 0,06$ e $0,34 \pm 0,005$, respectivamente, para a herdabilidade no sentido restrito para o caráter diâmetro em *Pinus caribaea* var. *hondurensis*, confirmado a grande precisão do método bayesiano.

Quanto à estimativa dos efeitos fixos (efeitos de blocos, por exemplo) e a predição dos efeitos aleatórios (valores genéticos), tem-se que as médias das distribuições marginais a posteriori dos parâmetros de locação (efeitos fixos e aleatórios), dados os componentes de variância ou parâmetros de dispersão, equivalem às soluções das equações do modelo misto do BLUP, desde que: sejam atribuídas prioris não informativas para os efeitos fixos, prioris normais para os efeitos aleatórios e verossimilhança normal para o vetor de observações. Neste caso, a vantagem da abordagem bayesiana refere-se à obtenção de valores genéticos preditos com menores desvios padrões.

Em resumo, a análise bayesiana propicia estimativas mais precisas de componentes de variância, parâmetros genéticos, valores genéticos e ganhos genéticos. Adicionalmente, a estimativa bayesiana permite a análise exata de amostras de tamanho finito. Este último aspecto é muito importante, especial-

mente para programas de melhoramento baseados em conjuntos de dados desbalanceados, onde tal abordagem propicia uma elegante análise de amostra finita, a qual não pode ser obtida pela metodologia freqüentista de modelos mistos.

12.2. Introdução

A predição de variáveis aleatórias e a estimativa de componentes de variância e efeitos fixos via metodologia de modelos mistos apresentam grande relevância na genética quantitativa aplicada ao melhoramento de plantas perenes. Embora a metodologia de modelos lineares mistos – MMLM – tratada pela abordagem da Inferência Estatística Freqüentista apresente várias propriedades desejáveis, publicações recentes têm mostrado uma superioridade da MMLM quando abordada do ponto de vista da Inferência Estatística Bayesiana.

Em 1986, o trabalho clássico de Gianola & Fernando (1986) propôs a abordagem bayesiana como uma estratégia conceitual para resolução de problemas em melhoramento animal, a qual foi detalhada por Gianola et al. (1990). Devido às excelentes propriedades teóricas e práticas da análise bayesiana, acredita-se que ela tornar-se-á rotineira em melhoramento de espécies perenes, transformando-se no procedimento padrão de estimativa de componentes de variância e de predição de valores genéticos.

Os geneticistas quantitativos foram influenciados pelo renascimento bayesiano que ocorreu em meados da década de 1960. Lindley & Smith (1972) apresentaram uma ligação entre modelos mistos e as abordagens bayesianas. Dempfle (1977) e Ronningen (1971) pesquisaram as conexões entre BLUP e estimadores bayesianos. Entretanto, somente na década de 1990, a utilização da Inferência bayesiana em melhoramento animal tornou-se rotineira, destacando-se os trabalhos de Gianola & Foulley (1990); Gianola et al. (1990), Weigel & Gianola (1993); Wang et al. (1993, 1994a e b); Jensen et al. (1994); Sorensen et al. (1994); Varona et al. (1994); Varona (1994); Van Tassel et al. (1995); Rodriguez et al. (1996); Van Arendonk et al. (1996), Garcia-Cortez & Sorensen (1996) e Rekaya (1997).

Outros trabalhos, enfatizando a estimativa bayesiana de parâmetros genéticos, mas não no contexto dos modelos lineares mistos, foram desenvolvidos (Toro & Prunonosa, 1984; Saio & Hayashi, 1990) através do uso do espaço paramétrico como informações a priori. No Brasil, a análise Bayesiana vem sendo recomendada para avaliação genética em melhoramento animal (Magnabosco, 1997; Rosa, 1999) e florestal (Resende, 1997; 1999b).

A estimativa bayesiana de variáveis aleatórias apresenta grande relevância em diversas áreas do conhecimento além da genética quantitativa, especialmente na biometria, econometria, engenharia, física e inteligência artificial, dentre outras. Segundo Sun et al. (1996), de maneira genérica, uma análise bayesiana exata pode ser obtida para os modelos de componentes de variância baseados tradicionalmente na teoria da normalidade, permitindo, para qualquer

parâmetro de interesse, uma detalhada inferência, para amostras de tamanho finito. Esta é uma das grandes vantagens da análise bayesiana.

12.3. Estimação e Predição Bayesiana

A origem da Inferência Estatística Bayesiana ocorreu com lançamento da obra *An Essay Towards Solving a Problem in the Doctrine of Chances*, a título póstumo, de autoria do reverendo Thomas Bayes. Esta obra foi publicada em 1763, por Richard Price, e apresenta o teorema denominado Teorema de Bayes (Murteira, 1990).

O teorema de Bayes fundamenta-se em probabilidades condicionais. Para enunciar tal teorema torna-se relevante apresentar alguns conceitos básicos de probabilidade, conceitos estes apresentados em vários livros de estatística, por exemplo, em Mood et al. (1974).

12.3.1. Conceito de probabilidade, probabilidade condicional e Teorema de Bayes

O conceito de probabilidade é a base de toda a Estatística. A definição de probabilidade é aceita por praticamente todos os estatísticos. Entretanto, várias interpretações de probabilidade existem, destacando-se a Física ou Clássica, a Freqüentista, a Lógica e a Subjetiva.

(i) Probabilidade Freqüentista

Sob o enfoque freqüentista, as freqüências relativas de ocorrência dos resultados convergem para a probabilidade de ocorrência destes resultados quando o número n de repetições do experimento tende a ser grande. Por exemplo, a freqüência de um resultado A é dada por $f_A = \frac{n_A}{n} \Rightarrow \lim_{n \rightarrow \infty} f_A = P[A]$, em que

n_A é o número de vezes em que ocorreu o resultado A e $P[A]$ denota a probabilidade de A .

(ii) Probabilidade Física ou Clássica

A interpretação clássica de probabilidade baseia-se no conceito de resultados igualmente prováveis. Assim, a probabilidade de um resultado ou evento é dada pela razão $1/N$, em que N é o número de resultados possíveis.

(iii) Probabilidade Subjetiva

Segundo esta interpretação, a probabilidade que uma pessoa atribui a um possível resultado de algum processo representa seu próprio julgamento da chance de o resultado ser obtido. Assim, a probabilidade de um evento A é a medida do grau de confiança que a pessoa deposita em A .

(iv) Probabilidade Lógica

A interpretação lógica de probabilidade pode ser derivada em função do conhecimento científico que é de domínio público, ou seja, do conhecimento comum a todos.

A seguir são apresentadas algumas definições relevantes de Probabilidade.

Espaço de probabilidade: um espaço de probabilidade é a tripla $(\Omega, A, P[\cdot])$, onde Ω é o espaço amostral, A é uma coleção de eventos e $P[\cdot]$ é uma função de probabilidade com domínio em A .

Probabilidade condicional: Se A e B são eventos em A , para um dado espaço de probabilidade $(\Omega, A, P[\cdot])$, a probabilidade condicional de um evento A dado o evento B , indicado por $P[A|B]$ é definida por:

$$P[A|B] = \frac{P[A,B]}{P[B]} \text{ se } P[B] > 0,$$

$$P[A,B] = P[B] \cdot P[A|B] = P[A] \cdot P[B|A].$$

Desta forma, como $P[A|B]$ é uma probabilidade, valem para ela todas as propriedades de probabilidade. A prova de que a proporção $P[A,B]/P[B]$ é uma probabilidade vem da verificação dos axiomas:

- (i) $P[A|B] = P[A,B]/P[B] \geq 0$
- (ii) $P[\Omega|B] = P[\Omega,B]/P[B] = P[B]/P[B] = 1$
- (iii) Se A_1, A_2, \dots, A_n ..., é uma seqüência de eventos mutuamente exclusivos em A , então:

$$P\left[\bigcup_{i=1}^{\infty} A_i|B\right] = \sum_{i=1}^{\infty} P[A_i|B]$$

Regra da multiplicação: Para um dado espaço de probabilidade $(\Omega, A, P[\cdot])$, se A_1, A_2, \dots, A_n são eventos em A , para o qual $[A_1, A_2, \dots, A_{n-1}] > 0$; então:

$$P[A_1 \cdot A_2 \cdot \dots \cdot A_n] = P[A_1] \cdot P[A_2 | A_1] \cdot P[A_3 | A_1 A_2] \cdot \dots \cdot P[A_n | A_1 \cdot A_2 \cdot \dots \cdot A_{n-1}]$$

Eventos independentes: Para um dado espaço de probabilidade $(\Omega, A, P[\cdot])$, sejam A e B dois eventos em A . Os eventos A e B são definidos como independentes se, e somente se, uma das condições for satisfeita:

- (i) $P[A,B] = P[A] \cdot P[B]$
- (ii) $P[A|B] = P[A]$ se $P[B] > 0$
- (iii) $P[B|A] = P[B]$ se $P[A] > 0$

Teorema de probabilidade total: Para um dado espaço de probabilidade $(\Omega, A, P[\cdot])$, se B_1, B_2, \dots, B_n é uma coleção de eventos mutuamente disjuntos em A satisfazendo:

$$\Omega = \bigcup_{j=1}^n B_j \text{ e } P[B_j] > 0 \text{ para } j=1, 2, \dots, n \text{ então para todo } A \in \mathcal{A}:$$

$$P[A] = \sum_{j=1}^n P[A|B_j] \cdot P[B_j] = P[A]$$

Considere o seguinte exercício: duas caixas possuem bolas vermelhas (V) e azuis (A). A caixa 1 (C_1) possui 60V e 40A enquanto a caixa 2 (C_2) possui 10V e 20A. Suponha que uma caixa seja tomada ao acaso e uma bola tomada aleatoriamente dentro desta caixa. Qual a probabilidade de que a bola seja vermelha?

Tem-se:

$$P[C_1] = 0,5; P[C_2] = 0,5$$

$$P[V|C_1] = 0,60; P[V|C_2] = 10/30 = 0,3333$$

Aplicando o teorema da probabilidade total, tem-se:

$$\begin{aligned} P[V] &= P[V \cap C_1] + P[V \cap C_2] \\ &= P[V|C_1] \cdot P[C_1] + P[V|C_2] \cdot P[C_2] \\ &= 0,60 \cdot 0,5 + 0,3333 \cdot 0,5 \\ &= 0,4666 \end{aligned}$$

Teorema de Bayes: Para um dado espaço de probabilidade $(\Omega, \mathcal{A}, P[\cdot])$, se B_1, B_2, \dots, B_n é uma coleção de eventos mutuamente disjuntos em \mathcal{A} satisfazendo:

$$\Omega = \bigcup_{j=1}^n B_j \text{ e } P[B_j] > 0 \text{ para } j=1, 2, \dots, n \text{ então para todo } A \in \mathcal{A}:$$

$$P[B_k|A] = \frac{P[A|B_k] \cdot P[B_k]}{\sum_{j=1}^n P[A|B_j] \cdot P[B_j]}.$$

Considere novamente o exemplo das bolas e caixas. Qual a probabilidade de ter sido tomada a caixa 1, sabendo-se que a bola amostrada foi a vermelha?

Aplicando-se o teorema de Bayes, tem-se:

$$P[C_1|V] = \frac{P[V|C_1] \cdot P[C_1]}{P[V|C_1] \cdot P[C_1] + P[V|C_2] \cdot P[C_2]} = \frac{P[V|C_1] \cdot P[C_1]}{P[V]} = \frac{0,60 \cdot 0,5}{0,4666} = 0,6428$$

Assim, o Teorema de Bayes envolve os conceitos de Probabilidade Condicional e Probabilidade Total.

O Teorema de Bayes, definido em termos de densidades de probabilidade, tem a seguinte formulação para a distribuição de uma variável aleatória contínua:

$$f(\theta|y) = \frac{f(y|\theta) f(\theta)}{\int_R f(y|\theta) f(\theta) d\theta}.$$

(1), onde:

θ : vetor de parâmetros

y : vetor de dados ou de informações obtidas por amostragem

$f(\theta|y)$: distribuição condicional de θ dado y , ou distribuição a posteriori (que é a base da estimação e predição bayesiana).

$f(y|\theta)$: função densidade de probabilidade da distribuição condicional de uma observação (y) dado θ (denominada função de verossimilhança ou modelo para os dados).

$f(\theta)$: função densidade de probabilidade da distribuição a priori, que é também a densidade marginal de θ . Esta função denota o grau de conhecimento acumulado sobre θ , antes da observação de y .

$f(y|\theta) f(\theta)$: função densidade conjunta de y e θ .

$f(y) = \int_R f(y, \theta) d\theta = \int_R f(y|\theta) f(\theta) d\theta = E_\theta[f(y|\theta)]$ - distribuição marginal ou preditiva de y com respeito a θ , onde R é a amplitude da distribuição de θ . E_θ significa esperança com respeito à distribuição de θ . (A integração da distribuição conjunta, no espaço paramétrico, produz a marginal de y).

Como $f(y)$ não é função de θ (ou seja, $f(y)$ é constante para qualquer θ), a forma usual da formulação de Bayes é: $f(\theta|y) \propto f(y|\theta) f(\theta)$, onde \propto indica proporcionalidade.

A expressão (1) advém das expressões $f(\theta,y) = f(y|\theta) f(\theta)$ e $f(\theta,y) = f(\theta|y) f(y)$, as quais são obtidas a partir do teorema da probabilidade condicional.

Verifica-se pela expressão (1) que um fator lógico que diferencia o enfoque bayesiano da abordagem freqüentista refere-se ao tipo de informação utilizada. Na concepção bayesiana, toda informação de que se dispõe é útil e deve ser utilizada. Por outro lado, a estatística clássica utiliza apenas observações de dados reais, desprezando as informações subjetivas (Gamerman & Migon, 1993).

O componente amostral ou experimental é comum aos modelos clássicos e bayesianos, mas com interpretações diferentes. Embora os modelos bayesianos passem por uma extensão dos modelos clássicos, existe uma divergência fundamental entre os dois enfoques: no modelo clássico o parâmetro é um escalar ou um vetor desconhecido, porém fixo, ao passo que no modelo bayesiano o parâmetro é considerado como escalar ou vetor aleatório (não observável), pois para os bayesianos tudo o que é desconhecido é incerto e, portanto, toda a incerteza deve ser quantificada em termos de probabilidade (Murteira, 1988; 1990). Em função disto, os modelos bayesianos tratam formalmente a informação a priori, através da distribuição de probabilidade (subjetiva ou lógica) a priori. As informações a priori e amostrais permitem a atualização periódica da distribuição de probabilidade a posteriori e, portanto, permitem modificar e atualizar as estimativas dos parâmetros.

O teorema de Bayes é uma regra para atualização de probabilidades e pode ser utilizado para cálculo de probabilidades a posteriori em mais de um

estágio. Considerando, ainda, o exemplo das duas caixas com bolas vermelhas (V) e azuis (A), qual a probabilidade de ter sido amostrada a caixa 1 (C_1), tendo-se amostradas 2 bolas vermelhas?

$$P[C_1|V_1 \cap V_2] = \frac{P[V_1 \cap V_2 | C_1]P[C_1]}{P[V_1 \cap V_2]} = \frac{P[V_2|C_1]P[C_1|V_1]}{P[V_2|C_1]P[C_1|V_1] + P[V_2|C_2]P[C_2|V_1]}$$

O raciocínio usado na expressão acima é de que, no cálculo da probabilidade a posteriori, após a retirada da segunda bola vermelha V_2 , a probabilidade a priori a ser utilizada refere-se à probabilidade a posteriori no estágio anterior (após a retirada da primeira bola vermelha V_1). Assim, tem-se: $P[V_2|C_1] = P[V|C_1] = 0,60$;

$$P[C_1|V_1] = P[C_1|V] = 0,6428, \text{ conforme calculado anteriormente}$$

$$P[V_2|C_2] = P[V|C_2] = 0,3333$$

$$P[C_2|V_1] = 1 - P[C_1|V] = 0,3572$$

$$P[C_1|V_1 \cap V_2] = \frac{0,60 \cdot 0,6428}{0,60 \cdot 0,6428 + 0,3333 \cdot 0,3572} = 0,7641$$

A ordem em que as informações são processadas pelo Teorema de Bayes é irrelevante, sendo que a posteriori resultante é sempre a mesma. Também, as observações podem ser processadas em um único grupo, uma a uma, ou divididas em subgrupos que o resultado final permanece o mesmo (Gamerman, 1996).

A expressão (1) fornece a regra de atualização de probabilidades sobre θ , partindo de $f(\theta)$ e chegando a $f(\theta|y)$. Assim, a distribuição a posteriori é proporcional a verossimilhança x priori, ou seja, a função de verossimilhança conecta a priori a posteriori usando para isto os dados experimentais (amostrais). Dessa forma, a distribuição a posteriori contempla o grau de conhecimento prévio sobre o parâmetro $[f(\theta)]$ e também as informações adicionais propiciadas pelo experimento $[f(y|\theta)]$. A distribuição $[f(y)]$ preditiva ou marginal de y , antes de observar y é útil para verificar a adequação da priori através das previsões que ela fornece para y . Após observar y é útil para testar todo o modelo (Gamerman & Migon, 1993).

Assumindo distribuição normal para a priori e para as observações, ou seja, $\theta \sim N(\mu, r^2)$ e $(Y|\theta) \sim N(\theta, \sigma^2)$ com σ^2 conhecido, a distribuição a posteriori de θ é também normal, ou seja $(\theta|Y=y) \sim N(\mu_l, r_l^2)$, onde:

$$\mu_l = \frac{r^{-2}\mu + \sigma^{-2}y}{r^{-2} + \sigma^{-2}}; \quad r_l^{-2} = r^{-2} + \sigma^{-2}$$

Assim, verifica-se que a precisão ou inverso da variância conota informação, ou seja, a relação $w = r^{-2}/(r^{-2} + \sigma^{-2})$, $w \in (0,1)$, mede a informação contida na priori em relação à informação total (priori + verossimilhança). Dessa forma,

pode-se reescrever $\mu_1 = w \mu + (1-w) y$, de forma que a média da posteriori equivale à média ponderada pela certeza na priori e na verossimilhança. Verifica-se, também, que a precisão (r_1^{-2}) da posteriori equivale à soma das precisões da priori e da verossimilhança.

Em termos de estimação, enquanto para a estatística freqüentista podem existir vários estimadores para um determinado parâmetro, para a estatística bayesiana existe, em princípio, um único estimador, o qual conduz a estimativas que maximizam a função densidade de probabilidade a posteriori. Assim, inferências sobre θ são realizadas a partir da densidade a posteriori através da expressão geral $p(\theta|y) = \int_R f(\theta|y) d\theta$, onde p denota probabilidade (Gianola & Fernando, 1986).

Ao nível do $-i$ -ésimo elemento do vetor θ , a esperança condicional de θ_i dado y é $\int_R \theta_i f(y|\theta) f(\theta) d\theta / \int_R f(y|\theta) f(\theta) d\theta$, o qual é o usual estimador bayesiano de θ_i .

O teorema de Bayes pode ser estendido para inferências sobre dados não observados (dados futuros, por exemplo) (Lindley, 1965; Gamerman & Migon, 1993). Sendo \tilde{y} um vetor de dados não observados, tem-se $f(\tilde{y}, \theta|y) = f(\tilde{y}|\theta, y) f(\theta|y)$, (2)

que é a função densidade de probabilidade conjunta ou função preditiva. A função densidade de probabilidade marginal ou função preditiva de \tilde{y} é dada por

$$\begin{aligned} f(\tilde{y}|y) &= \int_{R^\theta} f(\tilde{y}, \theta|y) d\theta \\ &= \int_{R^\theta} f(\tilde{y}|\theta, y) f(\theta|y) d\theta \\ &= \int_{R^\theta} f(\tilde{y}|\theta) f(\theta|y) d\theta \\ &= E_{\theta|y} [f(\tilde{y}|\theta)] \end{aligned} \quad (3)$$

As expressões (2) e (3) representam a base para os problemas de previsão bayesiana. Do exposto, torna-se claro que a estimativa trata da compreensão de um fenômeno (realidade) através de afirmações probabilísticas a respeito de quantidades não observáveis e a previsão trata de fazer afirmações probabilísticas acerca de quantidades a serem observadas no futuro.

12.3.2. Função de verossimilhança

 A função de verossimilhança formaliza a contribuição dos dados amostrais para o conhecimento sobre θ , visto que conecta a distribuição a priori à distribuição a posteriori.

A função de verossimilhança $[\ell(\theta; y)]$ de θ é a função que associa a cada θ , o valor $f(y|\theta)$. Dessa forma $\theta \rightarrow \ell(\theta; y) = f(y|\theta)$. A função $\ell(\theta; y)$ associa

(para um valor fixo de y) a probabilidade de ser observado y a cada valor de θ . Assim, quanto maior o valor de ℓ maiores são as probabilidades atribuídas pelo particular valor de θ considerado, ao valor fixado de y . Ao fixar um valor de y e variar os valores de θ observa-se a plausibilidade ou verossimilhança de cada um dos valores de θ . É interessante observar que: $\int_R f(y|\theta) d\theta = 1$ mas $\int_\theta \ell(\theta; y) d\theta = K \neq 1$, ou seja, a integral da função densidade de probabilidade equivale a 1, mas a função de verossimilhança não integra 1 (Gamerman & Migon, 1993).

É importante ressaltar que os termos probabilidade e verossimilhança são conceitos diferentes. No cômputo da verossimilhança, fixa-se a amostra ou conjunto de dados (y) e varia-se o parâmetro θ , procurando encontrar o parâmetro verossímil ou plausível com o conjunto de dados. Por outro lado, no cálculo de uma probabilidade utiliza-se uma distribuição com parâmetro θ conhecido e calcula-se a probabilidade de observar um determinado valor $y = y_0$. Em outras palavras, o problema da probabilidade é predizer a chance de ocorrer y sabendo θ e o problema da verossimilhança é fazer afirmações sobre θ com base no valor observado y .

No cálculo de probabilidades fixa-se θ (conhecido) e varia-se y , ao passo que no cálculo da verossimilhança fixa-se y e varia-se θ .

Como exemplo, considere a distribuição $y \sim \text{Binomial}(2, \theta)$, em que:

$2 = n$: número de repetições do experimento.

θ = parâmetro da binomial = probabilidade de sucesso.

$y = 0, 1, 2$: espaço amostral (y é o número de sucessos).

$\theta \in (0,1)$: espaço paramétrico ou domínio de variação do parâmetro θ .

A função de probabilidade equivale a $p(y|\theta) = \binom{2}{y} \theta^y (1-\theta)^{2-y}$, a qual equi-

vale a própria função de verossimilhança, $\ell(\theta; y)$, quando se observa $y = y_0$. Considerando os diferentes elementos do espaço amostral, são as seguintes as funções de verossimilhança:

$$(i) \quad \text{se } y = 0, \text{ então } \ell(\theta; y = 0) = \binom{2}{0} \theta^0 (1-\theta)^2 = (1-\theta)^2;$$

$$(ii) \quad \text{se } y = 1, \text{ então } \ell(\theta; y = 1) = \binom{2}{1} \theta^1 (1-\theta)^1 = 2\theta(1-\theta);$$

$$(iii) \quad \text{se } y = 2, \text{ então } \ell(\theta; y = 2) = \binom{2}{2} \theta^2 (1-\theta)^0 = \theta^2.$$

Tendo-se observado $y = 0$, variando-se θ no intervalo de 0 a 1 na função de verossimilhança $\ell(\theta; y = 0) = (1-\theta)^2$, verifica-se que o ponto de máximo desta função ocorre quando $\theta = 0$. Por outro lado, tendo-se observado $y = 1$, variando-se θ no intervalo de 0 a 1 na função $\ell(\theta; y = 1) = 2\theta(1-\theta)$ verifica-se

que o ponto de máximo ocorre quando $\theta = \frac{1}{2}$. Finalmente, tendo-se observado $y = 2$, verifica-se que o máximo de $\ell(\theta; y = 2) = \theta^2$ equivale a 1. Estas funções de verossimilhança encontram-se plotadas na Fig. 12.1.

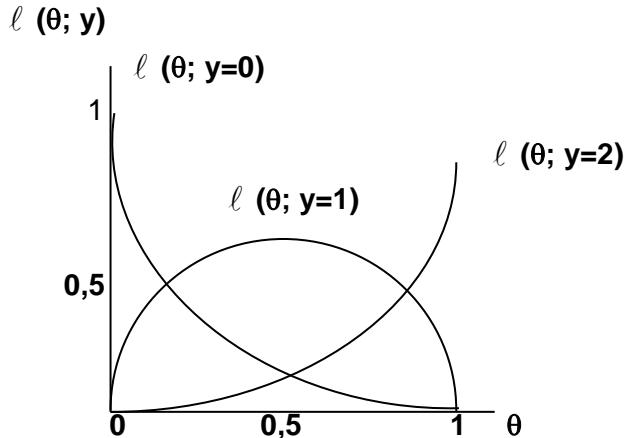


Fig. 12.1. Funções de verossimilhança para diferentes valores observados $y = 0, 1, 2$.

Em resumo, se for observado:

$$y = 0 \Rightarrow \hat{\theta} = 0 \text{ maximiza } \ell(\theta; 0);$$

$$y = 1 \Rightarrow \hat{\theta} = \frac{1}{2} \text{ maximiza } \ell(\theta; 1);$$

$$y = 2 \Rightarrow \hat{\theta} = 1 \text{ maximiza } \ell(\theta; 2).$$

Genericamente, uma função de verossimilhança pode ser construída conforme descrito a seguir. Sendo, y_1, y_2, \dots, y_n uma amostra aleatória independente e identicamente distribuída, com função de probabilidade ou função densidade de probabilidade dada por $f(y|\theta)$, em que θ é o parâmetro da função, a função de verossimilhança de $y = (y_1, y_2, \dots, y_n)$ é dada por $\ell(\theta; y) = f(y|\theta) = f(y_1, y_2, \dots, y_n|\theta)$, função essa que associa a cada θ , um valor de $f(y|\theta)$.

A função conjunta $\ell(\theta; y) = f(y|\theta) = f(y_1, y_2, \dots, y_n|\theta)$ fatora nas marginais de forma que equivale ao produtório $\ell(\theta; y) = f(y_1|\theta) \dots f(y_n|\theta) = \prod_{i=1}^n f(y_i|\theta)$. Como exemplo, considere a distribuição normal $N(\theta, \sigma^2)$. Neste caso, a função de verossimilhança desta normal equivale a $\ell(\theta; y) = f(y|\theta) = \prod_{i=1}^n f_N(y_i; \theta, \sigma^2) = \prod_{i=1}^n \frac{1}{[2\pi\sigma^2]^{1/2}} \exp\left\{-\frac{1}{2} \frac{(y_i - \theta)^2}{\sigma^2}\right\} \alpha \exp\left\{-\frac{n}{2\sigma^2} (\bar{y} - \theta)^2\right\}$, onde \bar{y} é a

$$\prod_{i=1}^n f_N(y_i; \theta, \sigma^2) = \prod_{i=1}^n \frac{1}{[2\pi\sigma^2]^{1/2}} \exp\left\{-\frac{1}{2} \frac{(y_i - \theta)^2}{\sigma^2}\right\} \alpha \exp\left\{-\frac{n}{2\sigma^2} (\bar{y} - \theta)^2\right\}, \text{ onde } \bar{y} \text{ é a}$$

média aritmética dos y_i .

A função de verossimilhança é a base do Princípio da Verossimilhança, o qual postula que toda a informação contida na amostra ou experimento encontra-se representada nesta função. Existem diferentes verossimilhanças que podem ser empregadas na estimação paramétrica, tais quais verossimilhança incondicional marginal, verossimilhança condicional e verossimilhança parcial. Gianola et al. (1989) apresentam detalhes sobre a estimação envolvendo estes diferentes conceitos.

12.3.3. Distribuições a priori

A distribuição a priori constitui-se no único elemento novo na análise bayesiana em relação à análise clássica. A determinação desta distribuição é subjetiva, entretanto dados experimentais podem ser utilizados nesta etapa. A premissa desta distribuição é que ela represente o estado atual de conhecimento sobre θ , antes de serem analisados os resultados experimentais. Zellner (1971) distingue dois tipos de informação a priori, uma baseada em dados passados e outra completamente subjetiva (não baseada em dados). Na presença de dados experimentais prévios ou quando baseada em fundamentos teóricos, a priori deve ser obtida de uma maneira científica (por formas funcionais).

A distribuição a priori tem importância maior quando a quantidade (n) de dados experimentais atuais é pequena. Caso contrário, a priori tende a ser dominada (pois recebe menor peso) pela verossimilhança, ou seja, com $n \rightarrow \infty$, as abordagens freqüentista e bayesiana são coincidentes. Em outras palavras, o peso da verossimilhança é maior do que o peso da priori, se a variância dos dados observados é menor que a variância da priori. É importante ressaltar, entretanto, que atribuir probabilidade zero a priori, significa atribuir zero também a posteriori, pois, neste caso, sendo $a_{posteriori} = priori \times verossimilhança$, não há verossimilhança que faça a posteriori diferente de zero.

12.3.4. Priori não informativa

Existem situações em que o conhecimento sobre determinado fenômeno é vago ou inexistente. Nestes casos, a distribuição a priori é dita vaga, difusa ou não informativa, não precisa ou de variância alta, significando que a densidade a priori reflete ignorância (Jeffreys, 1961).

Assim, a priori não informativa poderia ser representada por $f(\theta)\alpha$ constante; $-\infty < \theta < \infty$ (4)

Uma priori deste tipo poderia ser especificada através de uma distribuição com variância tendendo ao infinito (Gianola & Fernando, 1986) e, neste caso, maior ênfase ou peso serão dados à verossimilhança (dados experimentais ou amostrais).

Entretanto, para inferência sobre uma variância σ^2 , prioris da forma mencionada tornam-se informativas (Jeffreys, 1961) e uma priori não informativa

poderia ser representada por $f(\sigma^2) \propto \frac{1}{\sigma^2}$; $0 < \sigma^2 < \infty$.

Para prioris não informativas, supôs-se inicialmente distribuições uniformes como representantes de situações onde não se dispõe de informação inicial, ou seja, $f(\theta) \propto$ constante (θ variando na reta), fato que implica não favorecer qualquer valor particular de θ (Bayes, 1763).

Gamerman & Migon (1993) apresentam algumas dificuldades inerentes a esta escolha:

- (i) $f(\theta)$ é imprópria, ou seja, a integral sobre todos os possíveis valores de θ não converge: $\int f(\theta) d\theta \rightarrow \infty$.
- (ii) se $\phi = \phi(\theta)$ é transformação 1 a 1 (biunívoca) de θ e se θ tem distribuição uniforme, então, pelo teorema de transformações de variáveis, a densidade de ϕ será: $f(\phi) = f(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|$.

Assim, o raciocínio que conduz à especificação $f(\theta) \propto$ constante deveria levar também a $f(\phi) \propto$ constante, o que não é verdade. O ideal seria estabelecer uma regra que fosse invariante e que $f(\phi)$ não fosse imprópria.

Entretanto, o interesse principal reside na distribuição a posteriori e como esta é, em geral, própria (pois **(4)** é absorvida na constante de integração associada com a posteriori) mesmo quando a priori não o é, a eventual impropriedade das distribuições a priori não é importante (Gamerman & Migon, 1993). A classe de prioris não informativas proposta por Jeffreys (1961) é invariante, mas eventualmente imprópria.

12.3.5. Determinação subjetiva da priori

Sendo θ uma quantidade desconhecida e considerando os seus possíveis valores, a probabilidade a priori de cada valor possível de θ pode ser avaliada diretamente através de instrumentos auxiliares (loterias, roletas) quando θ for discreta. No caso de distribuições contínuas, têm sido sugeridas as abordagens do histograma, da função distribuição e da verossimilhança relativa (Gamerman & Migon, 1993).

12.3.6. Determinação da priori por formas funcionais

De algum conhecimento sobre θ pode-se descrever sua densidade por uma determinada forma funcional, ou eventualmente, pode-se definir uma família paramétrica de densidades. Adicionalmente deve-se verificar se a densidade escolhida está representando bem a informação disponível. Gamerman & Migon (1993) citam que, por exemplo, poder-se-ia fazer as seguintes considerações sobre θ :

- (i) θ distribui-se simetricamente em relação à moda;
- (ii) a densidade de θ decai rapidamente quando se afasta da moda, implicando uma variância pequena;
- (iii) intervalos muito afastados da moda tem probabilidade desprezível.

Tais considerações poderiam caracterizar aproximadamente a família da distribuição normal com seus hiperparâmetros, ou seja, parâmetros que auxiliam a especificação da priori.

Uma abordagem sistemática para a determinação de distribuições a priori pode ser enunciada tal qual o caso das distribuições conjugadas. A propriedade de preservação da classe de distribuições quando a distribuição do parâmetro é atualizada define conjugação.

No tópico 12.3.1., enunciou-se que se a distribuição dos dados é $(y|\theta) \sim N(\theta, \sigma^2)$ e a distribuição a priori é $\theta \sim N(u, r^2)$, a distribuição a posteriori de θ é $(\theta|y) \sim N(\mu_1, r_1^2)$. Assim, tem-se a priori e a posteriori com distribuição normal, ou seja, priori normal conjugada a posteriori normal. Este tipo de priori, segundo Gamerman & Migon (1993) torna a análise mais tratável permitindo explorar a natureza seqüencial do método bayesiano.

Sendo $F = \{f(y|\theta), \theta \in \Theta\}$ uma família de distribuições amostrais, uma classe de distribuição Ψ é conjugada a F se:

$$\forall f \in F \text{ e } f(\theta) \in \Psi \text{ então } f(\theta|y) \in \Psi$$

Quanto à definição de distribuições conjugadas, algumas ressalvas devem ser feitas (Gamerman & Migon, 1993):

- (i) a classe Ψ pode ser muito ampla, por exemplo, quando se tem $\Psi = \{\text{todas as distribuições}\}$ e F qualquer. Neste caso, a definição não tem valor prático;
- (ii) a classe Ψ pode ser muito restrita, por exemplo, quando $\Psi = \{P: P(\theta = \theta_0) = 1\}$. Isto implica que qualquer amostra não dará informação alguma, pois se tem a certeza de que $\theta = \theta_0$ e, portanto, esta certeza será mantida na posteriori. Assim, a definição também não terá valor prático. Dessa forma, recomenda-se então a atribuição de probabilidade positiva a todo e qualquer valor possível de θ por mais improvável que ele seja.

Em conclusão, a classe Ψ deve ser ampla o suficiente para proporcionar a escolha de uma distribuição a priori conveniente e, ao mesmo tempo deve ser restrita o suficiente para que a definição de distribuições conjugadas seja útil.

Como principais famílias conjugadas podem ser citadas:

- (i) família de distribuições Beta conjugada à Binomial.
- (ii) família de distribuições Normais conjugada à Normal.
- (iii) família de distribuições Gama conjugada à Poisson.
- (iv) família de distribuições Gama conjugada à Exponencial.
- (v) Normal-Gama conjugada à Normal.

12.3.7. Inferência bayesiana, inferência freqüentista e inferência verossimilhança

A inferência estatística tem evoluído muito nas últimas décadas, principalmente em função do aumento dos recursos computacionais. Desde o início do desenvolvimento da estatística até recentemente, houve grande supremacia da **inferência freqüentista**, a qual, em geral, tem sido creditada como fisheriana e se apóia, principalmente, na consistência assintótica através da qual, um modelo estimado converge para o verdadeiro modelo. A propriedade de incorporar a informação a priori na análise, sob o enfoque bayesiano, sempre gerou muita controvérsia, em determinada fase do desenvolvimento da estatística (Efron, 1986; Lindley, 1978; Smith, 1984). Atualmente, este tipo de discussão parece não ser mais relevante, com as diferenças entre as duas abordagens bem compreendidas e as virtudes de cada uma delas utilizadas quando mais conveniente (Gamerman, 1996).

Atualmente, uma das marcantes características da inferência estatística moderna é o papel central da função de verossimilhança, a qual foi introduzida por Fisher (1922). Segundo Lindsey (1999), poucos estatísticos modernos percebem que Fisher nunca foi um freqüentista para inferência (embora ele tenha geralmente usado uma interpretação freqüentista de probabilidade), sendo muito mais próximo aos bayesianos, tais quais Jeffreys, do que da escola de Neyman-Pearson (os verdadeiros freqüentistas, que introduziram a teoria dos testes de hipóteses entre 1936 e 1938). Segundo tal autor, a inferência estatística no sentido fisheriano enfatiza a obtenção do máximo de informação dado o conjunto de observações, sem incorporação de conhecimento e informação a priori, exceto aquelas necessárias para a construção de modelos (o princípio da verossimilhança baseia-se na suposição de que o modelo ou a função são verdadeiros e que somente os valores corretos dos parâmetros necessitam ser determinados). Portanto, a inferência Fisheriana refere-se, essencialmente, à **inferência verossimilhança**.

O uso da função de verossimilhança como um meio direto de fazer inferência, conforme sugerido por Fisher, é um enfoque recente em estatística (Edwards, 1972; Bardorff-Nielsen, 1976; Gomes, 1981; Pereira, 1997; Lindsey, 1999), caracterizando a **inferência verossimilhança**.

O artigo de Pereira (1997) caracteriza muito bem os pontos fortes da **inferência verossimilhança**, tais quais: (i) a função de verossimilhança fornece uma medida exata da incerteza, resumindo toda a informação que os dados fornecem sobre um parâmetro desconhecido θ ; (ii) a verossimilhança é uma medida relativa, de forma que não é possível obter uma medida de plausibilidade absoluta e, por isto, somente as razões de verossimilhança são relevantes; (iii) a função de verossimilhança relativa, dada por $R(\theta; y) = \frac{\theta^y(1-\theta)^{n-y} n^n}{y^n(n-y)^{n-y}}$ pode ser utilizada para obtenção de níveis de significância aproximados, através

dos testes da razão de verossimilhança; (iv) existe uma relação entre a função de verossimilhança e os métodos convencionais de teoria de grandes amostras, ou seja, além de fornecer uma medida exata de plausibilidade, a função de verossimilhança dá uma indicação de aplicabilidade de resultados assintóticos em casos particulares; (v) os métodos de verossimilhança são exatos tanto para pequenas quanto para grandes amostras, não dependendo de propriedades assintóticas das várias distribuições amostrais.

Assim, em **inferência bayesiana**, certos métodos que assumem distribuições a priori não informativas são essencialmente de **inferência verossimilhança**, tais como o método VEIL (ou da verossimilhança integrada de Gianola & Foulley, 1990) de estimação de componentes de variância, os quais mantêm a propriedade de conduzir a análise exata de amostra de tamanho finito.

12.4. Fundamentos Bayesianos da Predição de Valores Genéticos

Os fundamentos básicos da predição de valores genéticos são essencialmente de natureza bayesiana, conforme verificado inicialmente por Robertson (1955) e Dempfle (1977) e mostrado a seguir, considerando a distribuição a priori, normal, conjugada a posteriori normal.

Os efeitos genéticos aditivos são definidos como desvios e uma população de efeitos genéticos apresenta zero como média e σ_a^2 como variância. A melhor predição do efeito genético de um indivíduo sem nenhuma informação, tomado aleatoriamente da população é a média populacional μ , a qual pode ser tomada como o estimador a priori, cuja variância é σ_a^2 . Tomando uma informação y do indivíduo, um segundo (dado observado) estimador do efeito genético é o desvio ($y-\mu$) fenotípico em relação à média populacional, o qual possui variância $\sigma_e^2 = \sigma_y^2 - \sigma_a^2$. Estes dois estimadores independentes podem ser combinados linearmente da melhor maneira possível tomando as reciprocas das respectivas variâncias como pesos. Sob o enfoque bayesiano, a esperança da distribuição a posteriori corresponde à média ponderada pela precisão, das médias da priori e da verossimilhança (tópico 12.3.1.). Assim, tem-se:

$$\begin{aligned}\hat{a} &= \left(\frac{\text{estimador priori}}{\sigma_a^2} + \frac{\text{estimador informação observada}}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = \\ &= \left(\frac{0}{\sigma_a^2} + \frac{Y - \mu}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = h^2 (y - \mu),\end{aligned}$$

como no enfoque freqüentista, onde h^2 é a herdabilidade. Este é também (considerando y como a média dos indivíduos selecionados) o estimador do ganho genético com a seleção massal, conforme apresentado e derivado da mesma forma por Lerner (1958).

No melhoramento, duas fontes independentes de informação têm importante aplicação prática, mesmo quando se tem apenas o estimador a priori. Por exemplo, quando se deseja selecionar 30 indivíduos para reprodução, mas a partir de um teste de apenas 40 indivíduos tomadas aleatoriamente da população. Neste caso, o procedimento mais eficiente é selecionar os 20 melhores (que provavelmente estarão acima da média populacional) identificados com base no teste e tomar os outros 10 (não testados) aleatoriamente da população. Espera-se que dentre os 10 em torno de 5 estejam acima da média populacional. Robertson (1955) demonstrou claramente a utilidade de duas fontes de informação independentes para derivar várias fórmulas de predição.

12.5. Estimação Bayesiana de Índices de Seleção Unicaracterísticos

A predição de valores genéticos de genitores com base em testes de progênie é derivada a seguir, com base em Gianola & Fernando (1986), visando elucidar a relação entre índice de seleção e estimadores bayesianos.

Considere uma amostra de 5 genitores não parentados tomados aleatoriamente de uma população de genitores e suas capacidades de transmissão (metade de seus valores genéticos) representados pelo vetor $\theta' = [S_1, \dots, S_5]$, distribuídos a priori como $N(g, \sigma_S^2)$. Logicamente, neste caso, o interesse reside nas inferências sobre θ , dados os resultados do teste de progênie. O modelo para a média da progênie do genitor i é: $\bar{y}_i = \mu + S_i + \bar{e}_i$; $i=1, \dots, S$, onde μ é uma constante conhecida e $\bar{e}_i \sim N(0, \sigma_e^2/n_i)$ é a média dos resíduos associados com as observações da progênie n_i do genitor i . Uma vez que μ é considerado como conhecido, pode-se escrever: $w_i = \bar{y}_i - \mu = S_i + \bar{e}_i$.

$$E(w_i|S_i) = S_i \text{ e } \text{Var}(w_i|S_i) = \sigma_e^2/n_i.$$

Uma vez que os genitores não são parentados, a densidade a priori é dada por:

$$f(\theta) = \prod_{i=1}^S \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left\{-\frac{(S_i - g)^2}{2\sigma_S^2}\right\}$$

A função de verossimilhança, com $\text{Cov}(\bar{e}_i, \bar{e}_j) = 0$, $i \neq j$, é $f(w_1, \dots, w_S | \theta) =$

$$\prod_{i=1}^S \sqrt{\frac{n_i}{2\pi\sigma_e^2}} \exp\left\{-\frac{n_i(w_i - S_i)^2}{2\sigma_e^2}\right\}.$$

Com a suposição de independência

$$f(S_i|w_1, \dots, w_S) = f(S_i|w_i); i=1, \dots, S$$

A densidade a posteriori da capacidade de transmissão do i -ésimo genitor é então:

$$f(S_i|w_i) \propto \exp \left[-\frac{1}{2} \left(\frac{\sigma_s^2 + \sigma_e^2/n_i}{\sigma_s^2 \sigma_e^2/n_i} \right) \times \left(S_i - \frac{w_i \sigma_s^2 + g \sigma_e^2/n_i}{\sigma_s^2 + \sigma_e^2/n_i} \right)^2 \right].$$

Este resultado equivale a uma densidade normal com média:

$$E(S_i|w_i) = \frac{w_i \sigma_s^2 + g \sigma_e^2/n_i}{\sigma_s^2 + \sigma_e^2/n_i} = \frac{g (\sigma_s^2)^{-1} + w_i (\sigma_e^2/n_i)^{-1}}{(\sigma_s^2)^{-1} + (\sigma_e^2/n_i)^{-1}}.$$

Observa-se que a média a posteriori corresponde à média ponderada da média a priori (g) e da média da amostra (w_i), com os pesos sendo as recíprocas de suas variâncias.

Denominando $a = (\sigma_s^2)^{-1}/[(\sigma_s^2)^{-1} + (\sigma_e^2/n_i)^{-1}]$, tem-se:

$$\begin{aligned} E(S_i|w_i) &= ag + (1-a) w_i \\ &= g + (1-a)(w_i - g) = g + (1-a)(\bar{y}_i - \mu - g), \end{aligned}$$

que é a média condicional de uma distribuição normal bivariada.

Sendo $(1-a) = n_i/(n_i + \frac{\sigma_e^2}{\sigma_s^2})$, e sendo $g = 0$, neste caso tem-se:

$$E(S_i|w_i) = \frac{n_i}{n_i + \frac{\sigma_e^2}{\sigma_s^2}} (\bar{y}_i - \mu) = \frac{n_i}{n_i + (4-h^2)/h^2} (\bar{Y}_i - \mu),$$

que equivale ao clássico índice de seleção para avaliação da capacidade de transmissão de um touro via teste de progêneres (Van Vleck et al., 1987; Van Vleck, 1993a).

12.6. Relação entre BLUP e Estimadores Bayesianos

Duas derivações da metodologia BLUP sob o enfoque bayesiano são apresentadas a seguir, uma delas com base em Ronningen (1971) e Dempfle (1977, 1989) e outra com base em Robertson (1955).

Para ambas derivações tem-se o seguinte modelo e suposições, considerando G e R conhecidos:

$$y = X\beta + Za + e = Wt + e; W = [X Z]; t' = [\beta' a'];$$

$$E(a) = 0; E(e) = 0; E(y) = X\beta;$$

$$\text{Var}(a) = G; \text{var}(e) = R; \text{cov}(a, e) = 0; \text{var}(y) = ZGZ' + R.$$

A primeira derivação (Dempfle, 1977) usa o teorema de Bayes e considera-se que adicionalmente $t \sim N(r, M)$; $e \sim N(0, R)$; $y \sim N(X\beta, WMW' + R)$, onde:

$$E(t) = E \begin{pmatrix} \beta \\ a \end{pmatrix} = r = \begin{pmatrix} r_1 \\ 0 \end{pmatrix} \quad \text{e} \quad \text{var}(t) = M = \begin{pmatrix} S & 0 \\ 0 & G \end{pmatrix}$$

Dado $t, y \sim N(Wt, R)$, e dado $r, t \sim N(r, M)$. A verossimilhança de y dado t é $L(t; y) = k_1 \exp[-(y - Wt)'R^{-1}(y - Wt)/2] = f(y|t)$, e a densidade da distribuição a priori de t é $f(t) = k_2 \exp[-(t - r)'M^{-1}(t - r)/2]$.

Assim, a distribuição a posteriori de t é proporcional a

$$f(t|y) \propto L(t; y) f(t) \propto k_3 \exp\{-[(y - Wt)'R^{-1}(y - Wt) + (t - r)'M^{-1}(t - r)]/2\}.$$

Esta equação pode ser expressa, também, em termos da densidade conjunta a posteriori dada por:

$$f(a, \beta) \propto \exp\{-\frac{1}{2}(y - X\beta - Za)'R^{-1}(y - X\beta - Za)\}$$

$$\times \exp\{-\frac{1}{2}(\beta - r_1)'S^{-1}(\beta - r_1)\}$$

$$\times \exp\{-\frac{1}{2}(a - E(a))'G^{-1}(a - E(a))\}.$$

Uma vez que esta distribuição é simétrica e unimodal (normal), a moda, a mediana e a média são idênticas e uma grande classe de funções de perda comum (função de perda quadrática, função de perda absoluta ou função de perda uniforme) conduz ao mesmo estimador. Determinando a moda obtém-se o vetor médio da distribuição conjunta a posteriori, por maximização e não integração. Diferenciando a expressão com respeito a t (β e a) e igualando a zero obtém-se:

$$[W'R^{-1}W + M^{-1}] \hat{t} = W'R^{-1}y + M^{-1}r, \text{ onde } t \hat{t} = E[t|y].$$

Este sistema é equivalente a:

$$\begin{bmatrix} X'R^{-1}X + S^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} E(\beta|y) \\ E(a|y) \end{bmatrix} = \begin{bmatrix} X'R^{-1}y + S^{-1}r_1 \\ Z'R^{-1}y + G^{-1}0 \end{bmatrix} \quad (5)$$

em que $r_1 = E(\beta)$ e $0 = E(a)$.

Tomando a distribuição a priori sobre os efeitos fixos como não informativa (expressa como $S \rightarrow \infty$ e então $S^{-1} \rightarrow 0$), tem-se que esta equação resultante equivale às equações do modelo misto (EMM):

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

A equação para \hat{t} , pode ser derivada também pela maximização de $f(y, t)$ para variações em t , sendo o estimador, neste caso, denominado máximo a posteriori (MAP) (Henderson, 1982a, 1984).

A segunda derivação baseia-se em Robertson (1955) e fundamenta-se na combinação de dois estimadores (fontes de informação) independentes:

(i) o estimador BLUE ou de mínimos quadrados generalizados de t
 $\hat{t}_1 = (W'R^{-1}W)^{-1} W'R^{-1}y$ e

(ii) o estimador a priori de t

$$\hat{t}_2 = E(t) = r.$$

Estes estimadores apresentam as matrizes de variância-covariância

$$\text{var}(\hat{t}_1) = (W'R^{-1}W)^{-1} M \text{ e } \text{var}(\hat{t}_2) = 0, \text{ com variâncias do erro de estimação}$$

$$\text{var}(\hat{t}_1 - t) = (W'R^{-1}W)^{-1} \text{ e } \text{var}(\hat{t}_2 - t) = M$$

Combinando os dois estimadores obtém-se:

$$\hat{t}_c = [\text{var}(\hat{t}_1 - t)^{-1} + \text{var}(\hat{t}_2 - t)^{-1}]^{-1} \{\text{var}(\hat{t}_1 - t)^{-1} \hat{t}_1 + \text{var}(\hat{t}_2 - t)^{-1} \hat{t}_2\}, \text{ que equivale a } [WR^{-1}W + M^{-1}] \hat{t}_c = WR^{-1}y + M^{-1}r.$$

Fazendo as mesmas suposições sobre os efeitos fixos, esta expressão equivale às equações de modelo misto. As equações de modelo misto apresentadas em (5) são denominadas equações de modelo misto de Robertson.

12.7. Relação entre Estimadores de Máxima Verossimilhança e Estimadores Bayesianos

Considerando como uniforme a distribuição a priori dos parâmetros a serem estimados e maximizando (obtendo a moda) a distribuição a posteriori, o estimador resultante é equivalente ao de máxima verossimilhança – ML (Henderson, 1984; Gianola & Fernando, 1986). De fato, maximizando $f(a, \beta)$ (tópico anterior) (mas considerando uma priori não informativa para β) com respeito a a e β obtém-se um estimador denominado de máxima verossimilhança, por Henderson et al. (1959), embora $f(a, \beta)$ (tópico anterior) não seja uma função de verossimilhança e sim uma densidade a posteriori. Mesmo assim, pode ser obtido a partir das EMM que

$$E(\beta | y) = (X'V^{-1}X)^{-1} X'V^{-1}y = \hat{\beta} \text{ e}$$

$$E(a|y) = GZ'V^{-1} [y - X\beta] = \hat{a}$$

$\hat{\beta}$ é um estimador GLS e também ML de β e \hat{a} é um estimador ML de $E(a | \beta, y)$, equivalendo à média da distribuição condicional na qual β é fixado.

12.8. Estimação Bayesiana de Índices de Seleção Multicaracterísticos

Com a técnica dos índices de seleção multicaracterísticos deseja-se prever um vetor não observável $T = A\theta_1$ (denominado agregado genotípico) em

que A e θ_1 referem-se a uma matriz conhecida contendo os pesos econômicos relativos dos diferentes caracteres e um vetor de valores genéticos, respectivamente.

O vetor de dados é dado por:

$$y = \theta_2 + Z\theta_1 + e, \text{ onde:}$$

θ_2 : vetor de constantes especificando a locação da distribuição marginal de y

Z : matriz de incidência para θ_1

Para este modelo tem-se:

$$\begin{bmatrix} \theta_1 \\ e \\ y \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 & GZ \\ 0 & R & R \\ ZG & R & V \end{bmatrix} \right), \text{ onde } V = ZGZ' + R \text{ e } G \text{ e } R \text{ são conhecidas.}$$

A densidade a posteriori de θ_1 é $f(\theta_1|y) \propto f(y|\theta_1) \cdot f(\theta_1)$, a qual tem distribuição normal, pois $f(y|\theta_1)$ e $f(\theta_1)$ também tem distribuição normal. Esta posteriori equivale a:

$$\begin{aligned} f(\theta_1 | y) &\propto \exp \{ -\frac{1}{2} (y - \theta_2 - Z\theta_1)' R^{-1} (y - \theta_2 - Z\theta_1) \} \times \\ &\quad \times \exp \{ -\frac{1}{2} (\theta_1 - 0)' G^{-1} (\theta_1 - 0) \}, \text{ pois } E(\theta_1) = 0 \end{aligned}$$

Uma vez que esta distribuição é simétrica e unimodal, a média e a moda são a mesma, de forma que a distribuição pode ser diferenciada com respeito a θ_1 e igualada a zero para se obter a média a posteriori. Fazendo-se isto se obtém

$$Z'R^{-1}Z + G^{-1}E(\theta_1|y) = [Z'R^{-1}(y - \theta_2) + G^{-1}0]$$

Resolvendo-se esta equação para $E(\theta_1|y)$ obtém-se

$E(\theta_1|y) = (Z'R^{-1}Z + G^{-1})^{-1} [Z'R^{-1}(y - \theta_2) + G^{-1}0]$ que equivale à equação para o BLP ou índice de seleção associado a θ_1 . Dessa forma, necessita-se ainda incluir as informações econômicas para se obter T.

Uma vez que T é linear em θ_1 , a posteriori T é também normal com vetor médio $E(T|y) = A E(\theta_1|y)$.

Obtém-se, assim, o estimador bayesiano de T, bastando para isto, encontrar a média da distribuição a posteriori de $\theta_1|y$ e, posteriormente, incorporar as informações econômicas (A).

12.9. Estimação Bayesiana de Componentes de Variância e Relação com ML e REML

No contexto dos modelos lineares mistos, os valores genéticos (θ_1) são preditos simultaneamente à estimação dos efeitos fixos (θ_2) e dos componentes de variância (θ_3). Na abordagem bayesiana, a avaliação genética pode ser obti-

da, de maneira geral, pela construção da densidade a posteriori $f(\theta_1, \theta_2, \theta_3|y)$ e, se necessário, pela integração de $f(\theta_1, \theta_2, \theta_3|y)$ em relação a θ_2 e θ_3 . Estes (θ_2 e θ_3) são denominados parâmetros de *nuisance* e, por isso, devem ser integrados fora, exceto θ_1 em alguns casos, onde o mesmo constitui-se em uma parte integrante da função de mérito total (neste caso, a função de mérito depende da combinação linear de θ_1 e θ_2).

A obtenção de θ_1 requer a integração ou o conhecimento de θ_2 e θ_3 . Henderson (1973) propôs o método BLUP para situações em que θ_3 é conhecido e θ_2 não o é. Para situações em que θ_3 não é conhecido, este autor sugeriu que o procedimento de máxima verossimilhança (ML) propiciaria estimativas razoáveis. Conforme Gianola & Fernando (1986), argumentos bayesianos, que não requerem normalidade e linearidade, permitem validar a intuição de Henderson.

A distribuição de θ_1, θ_2 , e θ_3 , dado y é proporcional a

$$f(\theta_1, \theta_2, \theta_3|y) \propto f(y|\theta_1, \theta_2, \theta_3) \cdot f(\theta_1, \theta_2, \theta_3)$$

Concentrando o interesse em θ_1 (o vetor de valores genéticos), deve-se integrar θ_2 e θ_3 por meio de

$$f(\theta_1 | y) = \int_{R_{\theta_2}} \int_{R_{\theta_3}} f(\theta_1 | \theta_2, \theta_3, y) \cdot f(\theta_2, \theta_3|y) \cdot d\theta_2 d\theta_3.$$

Tomando a distribuição conjunta a posteriori de forma que a maioria da densidade esteja na moda ($\hat{\theta}_2, \hat{\theta}_3$), tem-se:

$$f(\theta_1 | y) \doteq f(\theta_1 | \theta_2 = \hat{\theta}_2, \theta_3 = \hat{\theta}_3, y).$$

Usando prioris não informativas para θ_2 e θ_3 , tem-se que $\hat{\theta}_2$ e $\hat{\theta}_3$ são precisamente estimadores ML de θ_2 e θ_3 , pois neste caso $f(\theta_2, \theta_3|y) \propto f(y|\theta_2, \theta_3)$, ou seja a densidade de θ_2 e θ_3 dado y é proporcional à função de verossimilhança, de forma que a moda da posteriori conjunta corresponde ao máximo da função de verossimilhança, produzindo estimadores ML.

Uma abordagem alternativa para inferência sobre θ_1 consiste em obter $f(\theta_1, \theta_2 | y) \doteq f(\theta_1, \theta_2 | \theta_3 = \hat{\theta}_3, y)$, onde $\hat{\theta}_3$ refere-se à moda da densidade marginal de θ_3 , dado y . Para obtenção de $\hat{\theta}_3$ deve-se integrar θ_2 em $f(\theta_2, \theta_3 | y) \propto f(y | \theta_2, \theta_3)$ e então maximizar $f(\theta_3|y)$. Usando-se uma priori não informativa para θ_3 , sob normalidade $\hat{\theta}_3$ é um estimador de máxima verossimilhança restrita (REML) para θ_3 (Harville, 1977). Assim, se o interesse reside na inferência conjunta para θ_1 e θ_2 basta usar $f(\theta_1, \theta_2 | y) \doteq f(\theta_1, \theta_2 | \theta_3 = \hat{\theta}_3, y)$, que sob normalidade é equivalente à solução das equações de modelo misto com θ_3 substituído pelas estimativas REML de θ_3 (desde que se tenha usado prioris não informativas para θ_2 e θ_3).

Inferências sobre componentes de variância devem ser baseadas em $f(\theta_3|y) \propto f(y|\theta_3) \cdot f(\theta_3)$, em que θ_3 contém variâncias e, portanto, $f(\theta_3|y)$ é definida na amplitude $(0, \infty)$ para cada um dos elementos de θ_3 , de forma que nunca surgem problemas de estimativas negativas de componentes de variância (Box & Tiao, 1973). $f(\theta_3|y)$ é obtida integrando-se θ_1 em $f(\theta_1, \theta_2, \theta_3|y)$ produzindo $f(\theta_2, \theta_3|y)$ e integrando-se θ_2 nesta última. Neste caso, $f(\theta_2, \theta_3|y)$ conduz aos estimadores ML de θ_2 e θ_3 e $f(\theta_3|y)$ conduz a um estimador REML de θ_3 . Segundo Giannola & Fernando (1986), isto (eliminação das influências de θ_2 ou dos efeitos fixos) mostra precisamente porque REML deve ser preferido em relação a ML, ou seja, estes argumentos são mais fortes do que os apresentados por Patterson & Thompson (1971), que enfatizaram a propriedade de vício do ML.

12.10. Estimadores Melhorados para a Média de Tratamentos

A análise bayesiana pode ser útil, também, no contexto da análise de experimentos de variados tipos e objetivos na pesquisa agropecuária.

Considerando o delineamento experimental de blocos ao acaso com t tratamentos, b blocos e n plantas por parcela (plantas obtidas aleatoriamente de uma população), tem-se o modelo matemático:

$Y_{ij} = \mu + t_i + b_j + e_{ij}$, onde μ , t_i , b_j e e_{ij} referem-se aos efeitos da média geral, de tratamentos, de blocos e do resíduo, respectivamente. Nas análises clássicas os efeitos t_i e e_{ij} podem ser considerados como aleatórios e b_j como fixo, visto que estas suposições não alteram os resultados da análise de variância, conforme Steel & Torrie (1980).

Aplicando o mesmo princípio (regressão) da predição de valores genéticos para a predição da média de tratamentos, tem-se que a fração de $(\bar{Y}_{..} - \bar{Y}_{..})$ (efeito de tratamento) que realmente determina a média do tratamento é $R^2(\bar{Y}_{..} - \bar{Y}_{..})$, em que:

$$R^2 = \sigma_t^2 / (\sigma_e^2 / b + \sigma_t^2) = 1 - 1/F, \text{ onde } F \text{ é o valor do teste F de Snedecor da análise de variância.}$$

Assim, a determinação (R^2) associada ao efeito de tratamento pode ser generalizada para efeitos fixos ou aleatórios de tratamentos, pois depende apenas de F que é idêntico para ambos os modelos. A média (valor esperado em uma nova repetição do ensaio ou em uma aplicação do tratamento ao nível operacional) verdadeira do tratamento (M) é então, melhor predita fazendo-se $M = (1 - 1/F)(\bar{Y}_{..} - \bar{Y}_{..}) + \bar{Y}_{..} = (1 - 1/F)\bar{Y}_{..} + (1/F)\bar{Y}_{..}$. Em outras palavras, a melhor predição da média é dada pela fração ($1 - 1/F$) da média aritmética do tratamento mais a fração ($1/F$) da média geral. Esta expressão para a média, além de ser mais coerente com a prática, considera adequadamente a precisão experimental na estimativa da média de tratamento, pois quanto mais preciso ($> F$) o expe-

rimento, menor é a correção ($-1/F$) aplicada sobre as médias de tratamento. Neste caso (agora supondo fixos os efeitos de tratamentos), níveis de significância diferentes (por exemplo 5% e 1%) conduzirão a médias ajustadas diferentes, embora na metodologia convencional conduzissem a inferências e médias idênticas.

Do ponto de vista da estatística freqüentista, a média amostral é o melhor estimador não viciado (BLUE) da média populacional. O estimador apresentado, que apresenta um regressor (no caso $1-1/F$) da média amostral do tratamento em relação à média geral é viciado (Gianola & Fernando, 1986). Entretanto, existem estimadores e preditores viciados que propiciam menor erro quadrático médio do que os estimadores BLUE (Henderson, 1984), sendo assim vantajosos em relação aos estimadores não viciados (Efron, 1975).

Os três principais procedimentos de estimação caracterizam os estimadores: melhor estimador linear não viciado (BLUE); estimador de mínimo erro quadrático médio e estimador bayesiano. Esses três estimadores são similares e idênticos sob certas circunstâncias (Henderson, 1984).

A utilização e o reconhecimento das vantagens dos estimadores viciados iniciaram-se com o trabalho de Stein (1955) que constituiu um verdadeiro paradoxo na Estatística. Stein demonstrou que a média aritmética é estimador não admissível, isto é, que existem estimadores que propiciam menor erro quadrático médio ou menor risco que a média aritmética. Neste contexto, James & Stein (1961), apresentaram um estimador melhorado para a média populacional, o qual é dado por $M^* = k(\bar{Y}_{..} - \bar{Y}_{..}) + \bar{Y}_{..}$, onde k é um fator regressor da média amostral de determinado tratamento sobre a média geral. Constatase assim, a similaridade entre M e M^* .

O fator k apresentado por James & Stein é dado por $K = 1 - \frac{(t-3) \sigma^2}{\sum (\bar{Y}_{..} - \bar{Y}_{..})^2}$, on-

de t refere-se ao número de tratamentos e σ^2 refere-se a variância residual.

Notando a similaridade entre $(t - 3)$ e o número de graus de liberdade $(t - 1)$ para tratamentos, verifica-se também a grande similaridade entre k e $(1 - 1/F)$. Stein (1955) demonstrou que a média aritmética é um estimador admissível apenas quando existem apenas uma ou duas médias a serem estimadas, de forma que a correção somente torna-se necessária quando três ou mais tratamentos forem considerados, conforme intuitivamente esperado. O estimador melhorado apresentado por James & Stein (1961), não necessita de qualquer suposição referente aos efeitos fixos ou aleatórios ou distribuições das médias a serem estimadas (Efron & Morris, 1977). Os estimadores de mínimo erro quadrático médio têm recebido mais atenção recentemente (Bibby & Toutenburg, 1977).

O procedimento de estimação bayesiana é mais antigo (1763) do que o método de Stein e também minimiza o erro quadrático esperado, de forma que o estimador de James-Stein é muito similar ao estimador de Bayes, tornando-se idênticos para grande número de tratamentos (Efron & Morris, 1977). Em inferência bayesiana não existe qualquer distinção entre efeitos fixos ou aleatórios,

sendo que os parâmetros a serem estimados são considerados variáveis aleatórias (Gianola & Fernando, 1986) que devem ser estimadas considerando as incertezas associadas a elas. No caso de inferências sobre médias populacionais de tratamentos, sob o enfoque bayesiano, Box & Tiao (1973) apresentam como regressor exatamente a quantidade $(1 - 1/F)$.

Com base no exposto, recomenda-se, para inferências (com menor risco) práticas na área de pesquisa agropecuária, a apresentação das médias de tratamento como $M = (1 - 1/F) \bar{Y}_{i..} + (1/F)\bar{Y}_{...}$, onde F , $\bar{Y}_{i..}$ e $\bar{Y}_{...}$ referem-se ao valor de F para tratamentos na análise de variância, média de tratamentos e média geral, respectivamente.

12.11. Problemas da Abordagem Clássica de Predição de Valores Genéticos e Vantagens da Abordagem Bayesiana

No procedimento de modelos mistos para avaliação genética, torna-se necessária a especificação de propriedades das distribuições dos vetores a e e , principalmente em termos de seus primeiro e segundo momentos. Para derivação do BLUP não é necessário considerar uma forma de distribuição, devendo-se assumir apenas que:

$$E(a) = 0; E(e) = 0; Var(a) = G; Var(e) = R; COV(a, e) = 0$$

Devido a isto se tem:

$$E(y) = X\beta; Var(y) = ZGZ' + R = V$$

Estas pressuposições são válidas em ausência de seleção. Em presença de seleção, $E(a)$ e $COV(a, e)$ podem ser diferentes de zero bem como G e R podem ser alteradas em relação a seus valores originais para uma população não selecionada e não endógamma.

Como propriedades desejáveis do BLUP, podem ser ressaltadas (Henderson, 1982a; Ronningen & Van Vleck, 1985):

- (i) é um preditor não viciado
- (ii) minimiza a variância do erro de predição dentro da classe das previsões lineares não viciadas
- (iii) minimiza o quadrado médio dos erros de predição dentro da classe das previsões lineares invariantes a translação (não afetadas por mudanças nos efeitos fixos)
- (iv) maximiza a correlação entre \hat{a}_i e a_i dentro da classe das previsões lineares não viciadas.

Considerando distribuição normal multivariada para a e e , propriedades adicionais interessantes são asseguradas (Henderson, 1982a):

- (i) Se os indivíduos candidatos à seleção apresentam a mesma quantidade de informação, a seleção maximiza a média esperada dos valores genéticos

- dos indivíduos selecionados $[E(a|y)]$. $(E(a|y))$ foi denominado *Best Predictor* por Henderson, 1973).
- (ii) \hat{a} é estimador ML de $E(a|y)$
 - (iii) \hat{a} é estimador MAP (máximo a posteriori) de a
 - (iv) $\hat{\beta}$ e \hat{a} são estimativas bayesianas, tomando-se como função de perda os desvios absolutos ou quadráticos e como estimativas a priori $E(a) = 0$ e $\text{Var}(\beta) = S$, com $S^{-1} \rightarrow 0$.
 - (v) As soluções BLUP \hat{a} e \hat{e} , podem ser usadas em algoritmos eficientes para estimação de componentes de variância por ML e REML, os quais permitem ao BLUP controlar vícios devidos a certos tipos de seleção.

Com relação aos efeitos da seleção na predição de valores genéticos, Henderson (1975, 1982b) mostrou (considerando normalidade) que, se os componentes de variância da população não selecionada são conhecidos, as equações de modelo misto ignorando a seleção ainda conduzem a predição BLUP sob seleção, desde que o critério de seleção seja função linear das observações e de translação invariante (ou seja, que o critério de seleção não envolva os efeitos fixos). Entretanto, predições envolvendo efeitos fixos baseadas em um modelo que ignora a seleção não são BLUP sob seleção. Fries & Schenkel (1993) abordam com detalhes esta questão.

Torna-se claro, então, que BLUP sob um modelo de seleção requer, pelo menos, o conhecimento das variâncias antes da seleção. Estimativas de componentes de variância pelos métodos ML e REML são livres de algumas formas de vício devido à seleção (Thompson, 1979; 1982; Gianola et al., 1989).

Considerar distribuição normal multivariada de y , a e e , implica considerar que os caracteres sejam determinados por um infinito número de locos não ligados de pequenos efeitos aditivos (modelo infinitesimal) (Mrode, 1996). Com o modelo infinitesimal, alterações na variância genética devidas à seleção, tais como desequilíbrio gamético (covariância negativa entre freqüências alélicas em diferentes locos) ou mudanças devidas à endogamia e deriva genética são consideradas nas equações de modelo misto através da inclusão da matriz de parentesco completa (Sorensen & Kennedy, 1983). Alterações na variância devidas a cruzamentos preferenciais, também são consideradas da mesma forma. Considerando o modelo infinitesimal, os componentes de variância podem ser estimados de forma não viciada por REML, desde que os dados utilizados incluam informações sobre as quais a seleção foi baseada.

Dentre as propriedades desejáveis do BLUP duas estão associadas à propriedade *Best*: minimização do quadrado médio do erro de predição e maximização da correlação entre os valores genéticos verdadeiros e os preditos. A terceira refere-se à propriedade de não vício $[E(a | \hat{a}) = \hat{a}]$. Estas três propriedades são interessantes do ponto de vista estatístico, entretanto, somente a propriedade de maximização da correlação entre a e \hat{a} apresenta uma conexão

direta com o problema do melhoramento por seleção. Uma vez que o objetivo das funções dos dados é ordenar os candidatos à seleção e praticar a seleção de forma a se maximizar o ganho genético com seleção, as propriedades de mínima variância e normalidade assintótica talvez sejam irrelevantes (Gianola & Fernando, 1986; Gianola, 1995). Esta assertiva pode ser confirmada considerando que com dados desbalanceados, independentemente da distribuição, o ordenamento dos candidatos com base em $E(a|y)$ e a seleção daqueles com os maiores valores, maximiza a média dos indivíduos selecionados, conforme demonstrado por Bulmer (1980); Goffinet (1983) e Fernando & Gianola (1986).

A seguir são relatados alguns pontos desfavoráveis da abordagem clássica e vantagens da inferência bayesiana para estimação. As propriedades desejáveis dos estimadores pontuais e intervalares na estatística clássica são baseadas em hipotéticas repetições do experimento em um número infinito de vezes e não consideram apenas os dados disponíveis. Isto significa que na teoria de amostragem, θ é uma quantidade conhecida fixa e o intervalo de confiança é aleatório. Assim, sob hipotética repetição de y , espera-se que o intervalo conteña θ em uma certa proporção das amostras (repetições). Entretanto, tendo-se os dados observados (uma repetição do vetor y), θ estará dentro ou fora do intervalo. Em análise bayesiana, inferências são realizadas diretamente da distribuição a posteriori, baseando-se apenas nos dados disponíveis e na distribuição a priori. A partir da posteriori, estimativas pontuais podem ser obtidas para resumir as características de tal distribuição e, também, estimativas intervalares e inferências probabilísticas podem ser realizadas. Por exemplo, no caso da seleção de genitores com base em testes de progêniens pode-se fazer inferências probabilísticas sobre a capacidade de transmissão do genitor i dado w_i , fazendo-se $Z = \frac{S_i - E(S_i | w_i)}{\sqrt{Var(S_i | w_i)}} \sim N(0,1)$ e computar para quaisquer valores fixados q_1 e q_2 , a probabilidade $P\{q_1 < S_i < q_2 | w_i\} = F(z_2) - F(z_1)$, onde F é a função distribuição acumulada da normal padrão.

Os procedimentos freqüentistas de estimação, ML e REML, apresentam propriedades desejáveis que são bem definidas somente para amostras muito grandes, ou seja, somente apresentam justificativas assintóticas. Entretanto, a maioria dos programas de melhoramento de espécies perenes baseia-se em amostras de tamanho finito (pequeno). O teorema de Bayes propicia soluções precisas para o problema de amostras de tamanho finito, pois para cada conjunto de dados, pequeno ou grande, existe uma distribuição a posteriori exata para realização de inferências. Existe, também, uma teoria assintótica bayesiana, de forma que com grandes amostras obtêm-se resultados similares aos obtidos com o método ML ou REML.

Em muitos casos, a distribuição dos dados (y) não depende exclusivamente do grupo de parâmetros de interesse (θ_1 , por exemplo), mas depende também de outros parâmetros incidentais denominados parâmetros de *nuisance* (θ_2 , por exemplo). Neste caso, a densidade conjunta a posteriori possui a forma $f(\theta_1,$

$\theta_2|y) \propto f(y|\theta_1, \theta_2) f(\theta_1, \theta_2)$, onde $f(\theta_1, \theta_2)$ é a densidade conjunta a priori de θ_1 e θ_2 . Uma vez que o interesse concentra-se em inferências sobre θ_1 na ausência de θ_2 , os parâmetros *nuisance* podem ser eliminados integrando-se a densidade conjunta a posteriori com respeito a θ_2 . Assim, tem-se:

$$\begin{aligned} f(\theta|y) &= \int_{R_{\theta_2}} f(\theta_1, \theta_2|y) d\theta_2 \\ &= \int_{R_{\theta_2}} f(\theta_1|\theta_2, y) f(\theta_2, y) d\theta_2 \end{aligned}$$

onde R_{θ_2} indica o espaço de θ_2 (Zellner, 1971).

Como exemplo, considere a questão da predição do valor genético de um indivíduo (θ_i) a partir de uma média fenotípica (y_i). A priori, $\theta_i \sim N.I.I.D. (0, h^2)$, $E(y_i|\theta_i) = \theta_i$ e $Var(y_i|\theta_i) = 1 - h^2$, onde h^2 é a herdabilidade. Neste caso, a densidade condicional $f(\theta_i|h^2 = K, y_i)$ é normal com média $E(\theta_i|h^2 = K, y_i) = Ky_i$.

Uma vez que existe apenas um dado (ponto), pode-se expressar $f(h^2|y_i) \propto f(h^2)$, onde $f(h^2)$ e $f(h^2|y_i)$ são as densidades a priori e marginal a posteriori da herdabilidade, respectivamente. Suponha que a priori h^2 pode ter três valores (0; 0,10 e 0,20) com probabilidades $P[h^2 = 0] = 0,10$; $P[h^2 = 0,10] = 0,40$; $P[h^2 = 0,20] = 0,50$. A herdabilidade neste exemplo é um parâmetro de *nuisance* uma vez que o interesse reside na inferência marginal sobre o valor genético. Assim, h^2 deve ser eliminado usando $f(\theta_i|y) = \int_{R_{\theta_2}} f(\theta_1|\theta_2, y) f(\theta_2|y) d\theta_2$ da seguinte forma:

$$\begin{aligned} f(\theta_i|y_i) &= f(\theta_i|h^2 = 0, y_i) \times 0,10 \\ &\quad + f(\theta_i|h^2 = 0,10, y_i) \times 0,40 \\ &\quad + f(\theta_i|h^2 = 0,20, y_i) \times 0,50 && \text{(usando o fato de que a} \\ &= 0,10 E(\theta_i|h^2 = 0, y_i) && \text{densidade condicional} \\ &\quad + 0,40 E(\theta_i|h^2 = 0,10, y_i) && f(\theta_i|h^2 = K, y_i) \text{ é normal} \\ &\quad + 0,50 E(\theta_i|h^2 = 0,20, y_i) && \text{com média} \\ &= 0,14 y_i && E(\theta_i|h^2 = K, y_i) = Ky_i). \end{aligned}$$

O valor 0,14 corresponde à média a priori da distribuição dos possíveis valores de herdabilidade. A expressão $f(\theta_i|y) = 0,14 y_i$ ilustra que a herdabilidade foi efetivamente integrada (eliminada), pois esta expressão não depende mais de h^2 (Gianola & Fernando, 1986).

Segundo Box & Tiao (1973) é muito importante examinar cuidadosamente $f(\theta_2|y)$. Se $f(\theta_2|y)$ é aguda, isto é, quando a maioria da densidade está próxima de sua moda ($\hat{\theta}_2$, diga-se), então $f(\theta_1|y) \doteq f(\theta_1|\theta_2 = \hat{\theta}_2, y)$ e as modas ou estimativas do tipo máxima verossimilhança (ML é modal pois refere-se a um ponto de máximo) podem ser usados como verdadeiros valores de $f(\theta_1|\theta_2, y)$. Como mostrado, isto é equivalente a realizar inferências com base na distribuição marginal

de θ_1 (valores genéticos) dado y (dados). Quanto a estimativas de componentes de variância, Gianola & Foulley (1990) mostraram que estimativas obtidas por REML são marginais apenas com respeito aos efeitos fixos mas são condicionais a outros parâmetros *nuisance* do modelo. A análise bayesiana propicia uma marginalização adicional que é particularmente interessante para modelos com grande número de componentes de variância.

Quanto à questão da predição de valores genéticos sob modelo de seleção, em análise bayesiana, as inferências sobre valores genéticos e parâmetros genéticos devem ser realizadas com base na densidade a posteriori construída como se a seleção e cruzamentos preferenciais não tivessem ocorrido. Os resultados obtidos dessa forma equivalem aos obtidos pela abordagem clássica considerando normalidade multivariada e são válidos para qualquer distribuição. Estes resultados são demonstrados a seguir com base em Gianola & Fernando (1986).

Considere θ_1 , um vetor desconhecido de valores genéticos, e θ_2 (efeitos fixos) e θ_3 (componentes de variância), vetores desconhecidos de parâmetros de *nuisance* e as decisões de seleção e cruzamento baseadas em dados disponíveis em y_0 (vetor $n_0 \times 1$). Dependendo de y_0 , dados adicionais y_i ($i = 1, \dots, s$) são obtidos a partir de decisões de seleção e cruzamento baseadas em y_0 . Em ausência de seleção e cruzamentos preferenciais $y_0, y_1, \dots, y_s, \theta_1, \theta_2$ e θ_3 tem distribuição conjunta com densidade $f(y_0, y_1, \dots, y_s, \theta_1, \theta_2, \theta_3)$.

A densidade marginal de $f(y_0, y_i, \theta_1, \theta_2, \theta_3)$ é obtida depois de apropriada integração e as inferências devem ser baseadas em $f(\theta_1, \theta_2, \theta_3 | y_0, y_i)$, com posterior integração (eliminação) de θ_2 e θ_3 .

Sendo R_i , a discreta e disjunta partição do espaço n_0 -dimensional, a densidade conjunta de $y_0, y_i, \theta_1, \theta_2, \theta_3$ é definida se $y_0 \in R_i$. Assim, pode-se escrever:

$$f(y_0, y_i, \theta_1, \theta_2, \theta_3 | y_0 \in R_i) = \frac{f(y_0, y_i, \theta_1, \theta_2, \theta_3)}{P[y_0 \in R_i]}$$

Usando probabilidade condicional, tem-se que:

$$f(y_0, y_i | y_0 \in R_i) = \frac{f(y_0, y_i)}{P[y_0 \in R_i]}$$

em que $f(y_0, y_i)$ é a densidade conjunta de y_0 e y_i , considerando que a seleção não tenha ocorrido. Combinando estas duas últimas expressões obtém-se:

$$\begin{aligned} f(\theta_1, \theta_2, \theta_3 | y_0, y_i, y_0 \in R_i) &= \frac{f(y_0, y_i, \theta_1, \theta_2, \theta_3) / P[y_0 \in R_i]}{f(y_0, y_i) / P[y_0 \in R_i]} \\ &= f(\theta_1, \theta_2, \theta_3 | y_0, y_i) \end{aligned}$$

Este resultado prova que se as decisões de seleção e cruzamento forem baseadas em y_0 e, se y_0 corresponde aos dados disponíveis para análise, as inferências sobre valores genéticos, efeitos fixos e componentes de variância devem ser realizadas a partir da densidade a posteriori considerando que a seleção não tenha ocorrido, concordando com os resultados de Henderson (1975, 1982b) e Thompson (1982) que são dependentes da suposição de normalidade. Assim, a distribuição condicional dado o conjunto completo de dados é a mesma distribuição em ausência ou presença de seleção e cruzamentos preferenciais e, para obter este resultado, a abordagem bayesiana não considerou linearidade, normalidade ou invariância à translação.

Em resumo, pode-se relatar as principais vantagens dos métodos bayesianos, conforme summarização realizada por Resende (1997):

- Rotineiramente componentes de variância são estimados pelo método REML e utilizados na predição de valores genéticos e cálculo do ganho genético. Entretanto, através deste procedimento, a distribuição e variância dos estimadores não são conhecidas e, assim, questões referentes à efetividade da seleção não podem ser respondidas com rigor. O método bayesiano de estimação de componentes de variância, implementado via simulação estocástica, como por exemplo através do procedimento denominado amostragem de Gibbs (GS), propicia uma descrição mais completa sobre a confiabilidade dos parâmetros genéticos do que o método REML. A abordagem Bayesiana (conhecimento da distribuição a posteriori dos parâmetros) possibilita a construção de intervalos de confiança (ou de probabilidade) exatos para as estimativas dos parâmetros genéticos, estando a confiabilidade (precisão) dos parâmetros associada com a quantidade e precisão das informações utilizadas. Dessa forma o método GS propicia estimativas mais precisas de componentes de variância, parâmetros genéticos, valores genéticos e ganhos genéticos. O método REML propicia apenas intervalos de confiança aproximados para os parâmetros genéticos, através do uso de aproximações e suposições de normalidade (argumentos assintóticos). De maneira geral, o método GS apresenta estimativas com menores desvios padrões que o método REML, especialmente para valores baixos de herdabilidade. A diferença de precisão entre os dois métodos reduz-se para conjuntos de dados extremamente grandes e, neste caso, tendem a ser obtidos resultados similares.
- Permite a utilização de informações a priori, a eliminação de parâmetros de *nuisance*, a integrada estimação – predição – decisão e a análise exata de amostras de tamanho finito (pois os dados são fixados na distribuição a posteriori). Este último aspecto é muito importante, especialmente para programas de melhoramento baseados em pequenos conjuntos de dados desbalanceados, onde a análise Bayesiana propicia uma elegante análise de amostra finita, a qual não pode ser obtida pela metodologia padrão de modelos mistos. O teorema de Bayes pro-

picia soluções precisas para o problema de amostras de tamanho finito, pois para cada conjunto de dados, pequeno ou grande, existe uma distribuição a posteriori exata para a realização de inferências. Existe, também, uma teoria assintótica Bayesiana, de forma que com grandes amostras se conseguem resultados similares aos obtidos com o método REML. Quanto aos parâmetros de *nuisance*, verifica-se que as estimativas de componentes de variância obtidas por REML são marginais apenas com respeito aos efeitos fixos, mas são condicionais a outros parâmetros *nuisance* do modelo. A análise Bayesiana permite uma marginalização adicional particularmente interessante para modelos com grande número de componentes de variância.

- A principal característica da análise Bayesiana refere-se à riqueza da informação propiciada por esta abordagem, possibilitando a obtenção de estimativas pontuais e os intervalos de probabilidade para as distribuições a posteriori dos parâmetros. Assim, inferências sobre parâmetros genéticos, valores genéticos e efeito de grupos genéticos podem ser realizadas com segurança.
- Sob modelo de seleção, as inferências sobre valores genéticos e parâmetros genéticos podem ser realizados com base na densidade a posteriori construída como se seleção e cruzamentos preferenciais não tivessem ocorrido, pois a distribuição destes parâmetros é a mesma com ou sem seleção e cruzamentos preferenciais (desde que sejam usados a matriz de parentesco completa e também os dados originais que orientaram as decisões de seleção).
- A análise bayesiana é capaz de resolver o problema dos valores genéticos ordenados, os quais apresentam distribuição desconhecida e cujas propriedades da distribuição normal não se aplicam.
- A abordagem Bayesiana apresenta, também, vantagens computacionais, pois não exige solução para as equações de modelo misto e, portanto, pode ser praticada em microcomputadores exigindo pouca memória (similar aos requerimentos de processos iterativos), embora o tempo de processamento seja muito maior.

12.12. Implementação Prática da Análise Bayesiana

Os resultados de interesse gerados pela análise Bayesiana são, em geral, as distribuições marginais a posteriori dos parâmetros genéticos, efeito de grupo genético e valores genéticos. Assim, inferências baseadas na média, mediana, moda e desvios padrões destas distribuições são realizadas na prática.

O problema básico da implementação da análise Bayesiana refere-se à integração numérica. A integração (no espaço do parâmetro) da função densidade de probabilidade a posteriori, por exemplo:

$$E [g (\theta)|y] = \int_{R_\theta} g (\theta) p(\theta|y) d\theta, \text{ onde:}$$

$g (\theta) = \theta$, para obtenção da média a posteriori e

$g (\theta) = (\theta - \mu)^2$, $\mu = E(\theta|y)$, para obtenção da variância a posteriori ou risco de Bayes, pode ser realizada através dos métodos (Gamerman, 1997): (i) analítico para aproximação de integral; (ii) automáticos ou de quadratura; (iii) simulação estocástica para obtenção de distribuições a posteriori. Os métodos analíticos para aproximação de integrais, tais como a Discrepância de Kulback Liebler, a Aproximação Assintótica à Normal e o Método de Laplace são funcionais até determinada dimensão do problema e estão associados a resultados assintóticos e à normalidade. O método de quadratura é puramente matemático e vem sendo substituído vantajosamente por métodos computacionalmente intensivos, os quais são essencialmente estatísticos, tais como os métodos de simulação estocástica. Dentre os métodos de simulação estocástica destacam-se: (i) Monte Carlo; (ii) Monte Carlo com Função de Importância; (iii) Reamostragem ou Bootstrap Bayesiano; (iv) Monte Carlo – Cadeias de Markov.

Dentre estas três grandes classes de algoritmos para aproximar as integrais, os métodos de Monte Carlo são largamente indicados e utilizados para integração multivariada. Os métodos de Monte Carlo referem-se a processos de aproximação de valores esperados (integrais com respeito a uma distribuição de probabilidade) por meio de amostras, podendo ser referidos também como um caso especial de simulação de um processo estocástico.

Em genética quantitativa, para implementação prática da análise Bayesiana, uma das maiores dificuldades técnicas é a marginalização. A obtenção de distribuições marginais por processos analíticos é praticamente impossível (Sorensen, 1996). Assim, a obtenção da distribuição marginal a posteriori (marginalização da distribuição conjunta a posteriori) tem sido obtida pelo método da amostragem de Gibbs (GS) através da amostragem e atualização das distribuições condicionais.

O método da amostragem de Gibbs pertence à classe de métodos, denominada Monte Carlo – Cadeias de Markov, a qual é sustentada em propriedades das Cadeias de Markov. O nome Gibbs advém da distribuição de Gibbs, que é muito utilizada na área de Física Estatística ou Mecânica Estatística. O amostrador de Gibbs explorando as distribuições condicionais completas através de algoritmo iterativo foi proposto inicialmente por Geman & Geman (1984) para aplicações na área de processamento de imagens. Entretanto, somente em 1990, este trabalho foi divulgado para toda a comunidade da área de estatística por Gelfand & Smith (1990) que publicaram em periódico da área de estatística, trabalho comparando o amostrador de Gibbs com outros processos de simulação estocástica. A partir de então, houve uma revolução na inferência estatística com uma impressionante realização de trabalhos baseados nesta técnica (Gamerman, 1996).

Para ilustrar a aplicação da técnica da amostragem de Gibbs na avaliação genética de espécies perenes será considerado o modelo individual univariado, conforme Sorensen (1996).

Modelo

$y = X\beta + Za + e$, onde:

y : vetor de dados, de ordem n .

β : vetor de efeitos fixos, de ordem p .

a : vetor de valores genéticos aditivos, de ordem q .

e : vetor de erros, de ordem n .

X, Z : matrizes de incidência que associam β e a aos dados (y).

Considera-se, inicialmente, que a distribuição condicional dos dados, dados β , a e σ_e^2 é normal multivariada: $y|\beta, a, \sigma_e^2 \sim N(X\beta + Za, I\sigma_e^2)$, onde I é a matriz identidade e σ_e^2 a variância residual.

Considerando o modelo quantitativo infinitesimal, tem-se que a distribuição de a é também normal multivariada: $a|A, \sigma_a^2 \sim N(O, A\sigma_a^2)$, onde A é a matriz de parentesco genético aditivo e σ_a^2 é a variância genética aditiva na população base.

Os parâmetros de interesse para inferências são: β , a , σ_a^2 e σ_e^2 . Para conduzir a análise Bayesiana torna-se necessário especificar as distribuições a priori para β , σ_a^2 e σ_e^2 (a distribuição de a já foi especificada).

Como priori para β pode-se assumir $p(\beta) \propto$ constante, que especifica aproximadamente a noção de conhecimento a priori vago para β . Esta distribuição a priori é imprópria, mas pode-se tornar própria, desde que se especifique os limites superior e inferior para $p(\beta)$.

As distribuições a priori dos componentes de variância (σ_e^2 e σ_a^2) poderiam ser uniforme da forma $p(\sigma_i^2) \propto$ constante, $0 \leq \sigma_i^2 < \sigma_{i \max}^2$ ($i = e, a$), onde, de acordo com o conhecimento acumulado sobre o caráter, $\sigma_{i \ max}^2$ seria o valor

máximo que σ_i^2 poderia assumir, a priori. Alternativamente, poderia ser especificada uma priori mais informativa para os componentes de variância, considerando uma distribuição qui-quadrado escalonada invertida, da forma: $p(\sigma_i^2 | \nu_i, S_i^2) \propto (\sigma_i^2)^{-(\nu_i/2+1)} \exp\left[\frac{-\nu_i S_i^2}{2\sigma_i^2}\right]$ ($i = e, a$), onde ν são os

graus de liberdade da distribuição qui-quadrado e S_i^2 , o valor inicial da variância. Esta distribuição reduz-se a uma distribuição uniforme imprópria se $\nu_i = -2$ e $S_i^2 = 0$.

Definidas estas distribuições, pode-se agora escrever a distribuição conjunta a posteriori dos parâmetros do modelo.

$$\begin{aligned} p(\beta, a, \sigma_a^2, \sigma_e^2 | y) &\propto p(\beta, a, \sigma_a^2, \sigma_e^2) p(y|\beta, a, \sigma_a^2, \sigma_e^2) \\ &= p(\beta) p(a|\sigma_a^2) p(\sigma_a^2) p(\sigma_e^2) p(y|\beta, a, \sigma_a^2, \sigma_e^2), \quad \text{em} \end{aligned}$$

que se omitiu o condicionamento nos hiperparâmetros (parâmetros que auxiliam na especificação da priori) e na conhecida matriz de parentesco A.

Considerando a distribuição a priori dos componentes de variância como uma qui-quadrado escalonada invertida, tem-se que a distribuição conjunta a posteriori pode ser reescrita:

$$\begin{aligned} p(\beta, a, \sigma_a^2, \sigma_e^2 | y) &\propto \sigma_e^{2\left(\frac{n+\nu_e}{2}+1\right)} \exp\left[-\frac{(y-X\beta-Za)'(y-X\beta-Za)+\nu_e S_e^2}{2\sigma_e^2}\right] \\ &\quad \sigma_a^{2\left(\frac{q+\nu_a}{2}+1\right)} \exp\left[-\frac{(a'A^{-1}a+\nu_a S_a^2)}{2\sigma_a^2}\right] \end{aligned}$$

Desejando atribuir distribuição a priori uniforme para σ_a^2 e σ_e^2 , basta fazer $\nu_i = -2$ e $S_i^2 = 0$ ($i=a, e$) na expressão acima.

Para implementação do GS, deve-se derivar todas as distribuições condicionais a posteriori a partir da distribuição conjunta a posteriori apresentada acima.

Denominando-se $X\beta+Za=W\theta$, onde $W=[X Z]$ e $\theta=[\beta' a']'$, tem-se que a matriz dos coeficientes das equações de modelo misto é dada por $C = W' W + \Sigma$, onde $\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & a^{-1}\sigma_e^2/\sigma_a^2 \end{bmatrix}$. A distribuição condicional a posteriori de θ é:

$\theta | \sigma_a^2, \sigma_e^2, y \sim N(\hat{\theta}, C^{-1}\sigma_e^2)$, em que $\hat{\theta}$ é dado por $C\hat{\theta}=W'y$, ou seja, pelas equações de modelo misto.

Para derivar a distribuição condicional a posteriori para β_i (o i-ésimo elemento do vetor β) devem ser feitas as partições:

$$\theta' = (\theta'_i, \theta'_{-i}); W' = (W_i W_{-i})'; C = \begin{bmatrix} C_{i,i} & C_{i,-i} \\ C_{-i,i} & C_{-i,-i} \end{bmatrix} = \begin{bmatrix} X'_i X_i & X'_i X_{-i} & X'_i Z \\ X'_{-i} X_i & X'_{-i} X_{-i} & X'_{-i} Z \\ Z' X_i & Z' X_{-i} & Z' Z + A^{-1}\alpha \end{bmatrix}, \quad \text{onde:}$$

$\alpha = \sigma_e^2 / \sigma_a^2$, β_{-i} é o vetor de efeitos fixos com o i-ésimo elemento excluído, de ordem $p-1$, X_i é o i-ésimo vetor coluna da matriz X e X_{-i} é a matriz X com X_i excluído.

Assim, tem-se: $\theta'_i | \theta'_{-i}, \sigma_a^2, \sigma_e^2, y \sim N(\hat{\theta}_i, C_{i,i}^{-1}\sigma_e^2)$, em que $\hat{\theta}_i$ é dado por $C_{i,i} \hat{\theta}_i = (W'_i y - C_{i,-i} \theta'_{-i})$.

Tem-se também:

$W_i y = X'_i y$; $(X'_i \ X_i) \hat{\beta}_i = X'_i y - X'_i \ X_{-i} \ \beta_{-i} - X'_i \ Z a$ e finalmente pode-se escrever a distribuição condicional a posteriori de β_i :

$$\beta_i | \beta_{-i}, a, \sigma_a^2, \sigma_e^2, y \sim N(\hat{\beta}_i, (X'_i \ X_i)^{-1} \sigma_e^2), \quad (6)$$

$$\text{em que: } \hat{\beta}_i = (X'_i \ X_i)^{-1} X'_i (y - X_{-i} \beta_{-i} - Z a)$$

Para derivar a distribuição condicional a posteriori de a_i , deve-se proceder de maneira similar. Assim, tem-se:

$$A^{-1} = \begin{bmatrix} A_{ii}^{-1} & A_{i,-i}^{-1} \\ A_{-i,i}^{-1} & A_{-i,-i}^{-1} \end{bmatrix}; \quad C = \begin{bmatrix} X' X & X' z_i & X' Z_{-i} \\ z'_i X & z'_i z_i + A_{i,i}^{-1} \alpha & z'_i Z_{-i} + A_{i,-i}^{-1} \alpha \\ Z'_{-i} X & Z'_{-i} z_i + A_{-i,i}^{-1} \alpha & Z'_{-i} Z_{-i} + A_{-i,-i}^{-1} \alpha \end{bmatrix}$$

$$\begin{aligned} W_i y = z'_i y; \quad (z'_i z_i + A_{i,i}^{-1} \alpha) \hat{a}_i &= z'_i y - z'_i X \beta - (z'_i Z_{-i} + A_{i,-i}^{-1} \alpha) a_{-i} \\ &= z'_i y - z'_i X \beta - A_{i,i}^{-1} \alpha a_{-i}, \quad \text{pois } z'_i Z_{-i} = 0 \end{aligned}$$

Finalmente, a distribuição condicional a posteriori de a_i é:

$$a_i | \beta, a_{-i}, \sigma_a^2, \sigma_e^2, y \sim N(\hat{a}_i, (z'_i z_i + A_{i,i}^{-1} \alpha)^{-1} \sigma_e^2) \quad (7)$$

A obtenção das distribuições condicionais a posteriori completas dos componentes de variância será descrita a seguir. Para σ_a^2 , deve-se tomar da distribuição conjunta a posteriori somente os termos que envolvem σ_a^2 . Assim:

$$p(\sigma_a^2 | \beta, a, \sigma_e^2, y) \propto (\sigma_a^2)^{-\left(\frac{q+\nu_a}{2}+1\right)} \exp\left[\frac{-a' A^{-1} a + \nu_a S_a^2}{2 \sigma_a^2}\right]$$

$$= (\sigma_a^2)^{-\left(\frac{\tilde{\nu}_a}{2}+1\right)} \exp\left[\frac{-\tilde{\nu}_a \tilde{S}_a^2}{2 \sigma_a^2}\right], \quad \text{em que}$$

$$\tilde{S}_a^2 = (a' A^{-1} a + \nu_a S_a^2) / \tilde{\nu}_a \quad \text{e } \tilde{\nu}_a = q + \nu_a.$$

A expressão de $p(\sigma_a^2 | \beta, a, \sigma_e^2, y)$ é proporcional à distribuição qui-quadrado (χ^2) escalonada invertida, com $\tilde{\nu}_a$ graus de liberdade e parâmetro de escala $a' A^{-1} a + \nu_a S_a^2$.

Assim:

$$\sigma_a^2 | \beta, a, \sigma_e^2, y \sim \tilde{v}_a \tilde{S}_a^2 \chi_{\tilde{v}_a}^2 \quad (8)$$

e para amostrar desta distribuição deve-se inicialmente trabalhar com um qui-quadrado com $\tilde{v}_a = q + v_a$ graus de liberdade, inverter este número e, posteriormente, multiplicá-lo por $a' A^{-1} a + v_a S_a^2$.

De maneira similar tem-se para σ_e^2 :

$$\begin{aligned} p(\sigma_e^2 | \beta, a, \sigma_a^2, y) &\propto (\sigma_e^2)^{-\left(\frac{n+v_e}{2}+1\right)} \exp\left[\frac{(y-X\beta-Za)'(y-X\beta-Za)+v_e S_e^2}{2 \sigma_e^2}\right] \\ &= (\sigma_e^2)^{-\left(\frac{\tilde{v}_e}{2}+1\right)} \exp\left[\frac{-\tilde{v}_e \tilde{S}_e^2}{2 \sigma_e^2}\right], \text{ onde} \end{aligned}$$

$$\tilde{v}_e = n + v_e \quad e \quad \tilde{S}_e^2 = ((y - X\beta - Za)'(y - X\beta - Za) + v_e S_e^2) / \tilde{v}_e.$$

A expressão de $p(\sigma_e^2 | \beta, a, \sigma_a^2, y)$ é proporcional à distribuição qui-quadrado escalonada invertida, com $\tilde{v}_e = n + v_e$ graus de liberdade e parâmetro de escala $(y - X\beta - Za)'(y - X\beta - Za) + v_e S_e^2$.

Assim, tem-se:

$$\sigma_e^2 | \beta, a, \sigma_a^2, y \sim \tilde{v}_e \tilde{S}_e^2 \chi_{\tilde{v}_e}^{-2} \quad (9)$$

A implementação da amostragem de Gibbs consiste em amostrar sucessivamente de (6), (7), (8) e (9).

Considere o exemplo:

Indivíduo	Bloco	Pai	Mãe	Y_i
1	1	-	-	-
2	2	-	-	-
3	1	1	-	Y_3
4	2	1	2	Y_4
5	1	3	4	Y_5
6	2	1	4	Y_6
7	1	5	6	-

Considerando o modelo individual tem-se $\beta = (\beta_1, \beta_2)'$ para os blocos 1 e 2 e $a = (a_1, \dots, a_7)'$ para os valores genéticos aditivos dos sete indivíduos. Considerando distribuições a priori uniforme para os componentes de variância, tais que $v_i = -2$, $S_i^2 = 0$, ($i = a, e$) e valores iniciais $\sigma_a^2 = 5,0$ e $\sigma_e^2 = 5,0$, deve-se trabalhar com as equações de modelo misto atualizando o valor $\alpha = \sigma_e^2 / \sigma_a^2$, a cada iteração. A matriz dos coeficientes das equações de modelo misto em função de α equivale a:

$$C = \begin{bmatrix} \beta_1 & \beta_2 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \\ \beta_1 & 2,00 & 0,00 & 0,00 & 1,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ \beta_2 & 0,00 & 2,00 & 0,00 & 0,00 & 1,00 & 0,00 & 1,00 & 0,00 \\ a_1 & 0,00 & 0,00 & 2,33\alpha & 0,50\alpha & -0,66\alpha & -0,50\alpha & 0,00\alpha & -1,00\alpha & 0,00\alpha \\ a_2 & 0,00 & 0,00 & 0,50\alpha & 1,50\alpha & 0,00\alpha & -1,00\alpha & 0,00\alpha & 0,00\alpha & 0,00\alpha \\ a_3 & 1,00 & 0,00 & -0,66\alpha & 0,00\alpha & 1+2,83\alpha & 0,50\alpha & -1,00\alpha & 0,00\alpha & 0,00\alpha \\ a_4 & 0,00 & 1,00 & -0,50\alpha & -1,00\alpha & 0,50\alpha & 1+3,00\alpha & -1,00\alpha & -1,00\alpha & 0,00\alpha \\ a_5 & 1,00 & 0,00 & 0,00\alpha & 0,00\alpha & -1,00\alpha & -1,00\alpha & 1+2,62\alpha & 0,62\alpha & -1,23\alpha \\ a_6 & 0,00 & 1,00 & -1,00\alpha & 0,00\alpha & 0,00\alpha & -1,00\alpha & 0,62\alpha & 1+2,62\alpha & -1,23\alpha \\ a_7 & 0,00 & 0,00 & 0,00\alpha & 0,00\alpha & 0,00\alpha & 0,00\alpha & -1,23\alpha & -1,23\alpha & 2,46\alpha \end{bmatrix}$$

Os parâmetros de locação são $\theta' = (\beta_1, \beta_2, a_1, \dots, a_7)$ e considerando que os valores iniciais são zero, tem-se $\theta^{(0)} = 0$ e pode-se iniciar a amostragem com uma cadeia de comprimento igual a n ($i = 1, \dots, n$).

A primeira amostragem para $\beta_1^{(1)}$ é obtida de (6) ou seja obtém-se $\beta_1^{(1)}$ de $\beta_1 | \beta_2, a, \sigma_a^2, \sigma_e^2, y \sim N(\hat{\beta}_1, (2)^{-1} \sigma_e^2)$. Dados os valores iniciais, a média da distribuição normal é $\hat{\beta}_1 = (Y_3 + Y_5)/2$ e a variância é $(2)^{-1} \sigma_e^2$, ou seja, 5/2.

$\beta_2^{(1)}$ por sua vez é retirado de uma distribuição normal com média $\hat{\beta}_2 = (Y_4 + Y_6)/2$ e variância 5/2.

$a_1^{(1)}$ por sua vez é retirado de $a_1 | \beta, a_{-1}, \sigma_a^2, \sigma_e^2, y \sim N(\hat{a}_1, (2,33\alpha)^{-1} \sigma_e^2)$, que provém de (7), onde $\hat{a}_1 = 0/2,33\alpha = 0$, $\sigma_e^2 = 5$ e $\alpha = 1$.

$a_2^{(1)}$ é retirado de uma distribuição normal $N(\hat{a}_2, (1,50\alpha)^{-1} \sigma_e^2)$, onde $\hat{a}_2 = (0 - 0,50\alpha a_1^{(1)})/(1,50\alpha)$, $\sigma_e^2 = 5$ e $\alpha = 1$, onde 0,50 provém de C.

Para o indivíduo número 6, o valor genético na primeira iteração, $a_6^{(1)}$, é retirado de:

$N(\hat{a}_6, (1+2,62\alpha)^{-1} \sigma_e^2)$, onde

$$\hat{a}_6 = \frac{Y_6 - \beta_2^{(1)} - (-1,00\alpha a_1^{(1)} - 1,00\alpha a_4^{(1)} + 0,62\alpha a_5^{(1)})}{1+2,62\alpha}, \sigma_e^2 = 5 \text{ e } \alpha = 1.$$

Para o indivíduo número 7, o valor genético na primeira iteração, $\hat{a}_7^{(1)}$, é retirado de:

$N(\hat{a}_7, (2,46\alpha)^{-1}\sigma_e^2)$, onde

$$\hat{a}_7 = \frac{0 - (-1,23\alpha a_5^{(1)} - 1,23\alpha a_6^{(1)})}{2,46\alpha}, \quad \sigma_e^2 = 5 \text{ e } \alpha = 1.$$

Tendo amostrado todos os parâmetros de locação do modelo, deve-se computar:

$$SS_e^{(1)} = (y - X\beta^{(1)} - Za^{(1)})'(y - X\beta^{(1)} - Za^{(1)})$$

$$SS_a^{(1)} = (a^{(1)})' A^{-1} a^{(1)}$$

A primeira iteração do amostrador é completada, retirando-se os componentes de variância, usando $SS_a^{(1)}$ e $SS_e^{(1)}$:

$$\sigma_a^2 | \beta, a, \sigma_e^2, y \sim SS_a^{(1)} \chi_{q-2}^{-2} \quad (10)$$

$$\sigma_e^2 | \beta, a, \sigma_a^2, y \sim SS_e^{(1)} \chi_{n-2}^{-2} \quad (11)$$

A segunda iteração inicia-se através de atualizações das equações de modelo misto com $\alpha = \sigma_e^2 / \sigma_a^2$, onde σ_e^2 e σ_a^2 são os valores amostrados de (10) e (11) e prossegue com a amostragem de $\theta^{(2)}$ da mesma maneira realizada para $\theta^{(1)}$ e com o cômputo de $SS_e^{(2)}$ e $SS_a^{(2)}$ da mesma maneira realizada para $SS_e^{(1)}$ e $SS_a^{(1)}$, e assim sucessivamente.

Em termos mais simples, o algoritmo GS pode ser apresentado de forma resumida:

1. Fornecer os valores iniciais dos parâmetros de locação e dispersão do modelo. Estes valores iniciais podem ser calculados através de procedimentos padrões tais como a estimativa de componentes de variância por REML ou quadrados mínimos. Considerando a média geral \bar{y} como único efeito fixo, pode-se calcular \bar{y} como a média aritmética das observações e $a_i = h^2(y_i - \bar{y})$. Devem ser fornecidos os valores iniciais para \bar{y}_i , a_i , σ_e^2 , σ_a^2 e $\alpha = \sigma_e^2 / \sigma_a^2$.
2. Gerar valores para os efeitos fixos. Sendo o único efeito fixo, a média geral, tem-se: $\hat{\bar{y}} = \bar{y} + md \sigma_e / (n)^{1/2}$
3. Gerar valores para os efeitos aleatórios: $\hat{a} = a_i + rnd [(1 - r_{IA}^2) \sigma_a^2]^{1/2}$, onde r_{IA} é a acurácia dada por $r_{IA} = (1 - PEV_i / \sigma_a^2)^{1/2}$, onde PEV_i é o ié-simo elemento da inversa da matriz dos coeficientes das EMM multiplicado por σ_e^2 .

4. Calcular a soma de quadrados do resíduo (SSE) e a variância residual σ_e^2 .

Considerando que a distribuição a priori para a variância residual é a inversa de uma qui-quadrado, tem-se:

$$SSE = \sum (y_i - \hat{y} - \hat{a}_i)^2$$

$$\sigma_e^2 = \frac{SSE}{X_n^2}$$

5. Gerar um valor para a variância dos efeitos aleatórios de valores genéticos.

$$\sigma_a^2 = \frac{\hat{a}' A^{-1} \hat{a}}{X_q^2}$$

6. Calcular o novo valor do parâmetro

$$\hat{\alpha} = \frac{\sigma_e^2}{\sigma_a^2}$$

7. Repetir os passos de (2) a (6) até que se obtenha a convergência da cadeia.

12.13. Diagnóstico de Convergência

Este tema foi apresentado em detalhes por Gamerman (1996) e aqui serão apresentados alguns aspectos práticos principais, com base na referida literatura.

As cadeias de Markov estão inseridas no contexto da teoria dos processos estocásticos, teoria esta, definida como a parte dinâmica da teoria de probabilidades, onde se estuda uma coleção de variáveis aleatórias, com respeito a sua interdependência e comportamento limite. Para a inferência bayesiana, é de maior relevância o estudo do comportamento assintótico da cadeia, quando o número de iterações tende a ∞ , uma vez que a inferência deve ser baseada na distribuição (a posteriori) estacionária, ou seja, em equilíbrio.

À medida em que o número de iterações aumenta, a cadeia se aproxima da condição de equilíbrio. Dessa forma, é necessário considerar a convergência em uma determinada iteração cuja distribuição esteja próxima da distribuição em equilíbrio (atingido teoricamente quando $n \rightarrow \infty$), ou seja, após um número suficientemente grande de iterações.

A forma básica de obter uma amostra de tamanho m da posteriori é produzir m cadeias independentes (geradas a partir de m valores iniciais diferentes) e, após a convergência, retirar os valores da última iteração de cada cadeia. Outra opção consiste em retirar m amostras da mesma cadeia após a convergência, visto que se estará amostrando da distribuição a posteriori em equilíbrio.

Neste último caso, é importante relatar que as amostras sucessivas não são independentes, de forma que se torna necessário descartar várias iterações entre cada duas amostras a serem salvas. Como o processo é markoviano, a dependência diminui com o aumento da distância entre iterações, obtendo-se, assim, independência entre as amostras salvas.

Considerando a segunda opção, no contexto do diagnóstico da convergência, tornam-se relevantes as quantidades: M = número de iterações pré-convergência ou período de descarte ou período de aquecimento da cadeia; N = número de iterações após a convergência; K = número de iterações entre amostras sucessivas ou intervalo entre amostras. O tamanho total da cadeia é dado por $T = M + N$.

O valor de K pode ser determinado calculando as autocorrelações na série de valores gerados e verificando a partir de qual ponto pode-se considerar as autocorrelações como nulas. Uma vez que o valor de K é muito menor que o de M , os métodos baseados em uma única cadeia (mais longa) são preferidos computacionalmente.

Outra forma de análise de convergência refere-se à estimativa do erro de Monte Carlo, conforme apresentado no tópico 12.15. Em resumo, para a aplicação prática, algumas recomendações são apresentadas a seguir.

Devido ao fato de que valores aleatórios são utilizados inicialmente como realização do conjunto de parâmetros, é necessário um período de descarte de amostras até que as amostras de GS possam ser consideradas como provenientes da distribuição conjunta a posteriori, ou seja, da distribuição em equilíbrio estacionário. Em geral, tem sido utilizado o esquema tradicional de cadeia longa (única) de Gibbs, onde o processo de reamostragem é contínuo. Assim, de maneira geral, um grande (da ordem de 10.000 a 1.000.000) número de ciclos tem sido utilizado, sendo descartadas as primeiras amostras (da ordem de poucos milhares) e amostras de cada parâmetro são salvas a cada pequeno (da ordem de 50 a 100) número de iterações. O intervalo entre amostras salvas é necessário como forma de obtenção de amostras independentes, visto que amostras sucessivas apresentam correlação serial. O número total de amostras salvas é utilizado para cômputo das estimativas pontuais e intervalares de interesse. O software MTGSAM (Van Tassell & Van Vleck, 1995) tem sido utilizado com sucesso para a implementação da análise bayesiana.

12.14. Estimação Bayesiana via Modelos Lineares Robustos

No tópico 12.12 foram consideradas verossimilhanças normais para os dados observados. Entretanto, a distribuição gaussiana usualmente considerada nos modelos de genética quantitativa nem sempre se ajusta aos dados observados. Neste caso, a presença de dados atípicos (outliers) pode distorcer os resultados e inferências obtidos com base na suposição de normalidade, visto que a distribuição normal não é robusta a observações discrepantes, conforme relato de Rogers & Tukey (1972).

Visando contornar este problema, distribuições simétricas de caudas longas (tais como a t de Student, a Slash e a normal contaminada) têm sido aplicadas recentemente (Strandén, 1996; Rosa, 1999) visando a estimativa robusta. Estas distribuições mencionadas pertencem ao grupo de distribuições denominado normal/independente. Conforme relatado por Rosa (1999), a distribuição normal/independente pode ser definida como a distribuição de um vetor k dimensional dado por $y = \mu + e/(w)^{1/2}$, onde μ é o vetor de locação de dimensão k , w é uma variável aleatória positiva com densidade $p(w|v)$, onde v é um parâmetro de robustez, e é um vetor aleatório k dimensional normalmente distribuído com média 0 e matriz de covariância Σ não singular.

Neste caso, a distribuição de y , condicional a w , é normal com média μ e matriz de covariância $\Sigma|w$. A distribuição de y é dada por:

$$p(y|\mu, \Sigma, v) = \int p(y|\mu, \Sigma, w) p(w|v) dw$$

Dessa forma, a distribuição de w define a distribuição de y . As distribuições normal/independente têm algumas propriedades comuns tais quais (Lange & Sinsheimer, 1993): (i) se $y \sim p(y|\mu, \Sigma, v)$, ou seja, se y tem distribuição normal/independente, então $E[y|\mu, \Sigma, v] = \mu$ e $Var[y|\mu, \Sigma, v] = E[w^{-1}|v] \Sigma$, desde que existam o primeiro e segundo momentos; (ii) $p(y|\mu, \Sigma, v)$ é simétrica e unimodal, de forma que a moda e a mediana são iguais à média μ .

Dentre as distribuições robustas mencionadas, a t de Student tem um único parâmetro de robustez, que são os graus de liberdade da distribuição. Quanto menor o número de graus de liberdade, mais pesadas são as caudas das distribuições t. Na distribuição Slash, a robustez também é definida por um único parâmetro. Por outro lado, a distribuição normal contaminada apresenta dois parâmetros de robustez, sendo que um fornece a proporção de contaminantes e o outro define como a variância das duas (contaminantes e não contaminantes) populações difere (Rosa, 1999). Verifica-se que estas distribuições robustas são modificações da distribuição normal (gaussiana), modificações estas em função dos parâmetros de robustez. Assim, a distribuição gaussiana é um caso particular de distribuição normal/independente equivalendo às distribuições t e Slash com parâmetro v infinito ou à distribuição normal contaminada quando a proporção de contaminantes é nula.

A implementação da análise Bayesiana usando distribuições normal/independente ao invés da suposição gaussiana pode ser realizada via amostrador de Gibbs, de maneira similar à descrita no tópico 12.12.

Em melhoramento animal, a presença de dados discrepantes pode ser devida a tratamento preferencial aos melhores animais, conforme constatado por Strandén (1996). Neste caso, os modelos robustos tendem a fornecer resultados mais plausíveis e devem ser adotados.

12.15. Aplicação a Dados Experimentais

A seguir são apresentados resultados obtidos por Resende (1999b) e Resende et al. (2001b), relativos à avaliação genética de um caráter quantitativo em uma espécie vegetal perene.

Foram utilizados dados de quatro experimentos (testes de progêneres de irmãos germanos) desbalanceados de *Pinus caribaea* var. *hondurensis* com um total de 89 progêneres, sendo 65 diferentes e algumas comuns aos experimentos. As 65 progêneres foram obtidas de 76 genitores e, no total, foram avaliados 3.886 indivíduos. Os experimentos foram implantados pela Champion Papel e Celulose S.A., no município de Santana no Estado do Amapá, no delineamento em blocos casualizados com nove repetições e cinco plantas por parcela. Considerou-se a variável diâmetro avaliada aos nove anos e meio e o modelo aditivo infinitesimal.

O modelo linear univariado equivale a: $y = X\beta + Za + Wc + e$, em que: y , β , a , c e e : vetores de dados, de efeitos fixos (blocos), de valores genéticos, de efeitos de ambiente comum (parcela) e erros aleatórios, respectivamente.

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

As seguintes distribuições foram consideradas:

$$y|\beta, a, c, \sigma_e^2 \sim N(X\beta + Za + Wc, I\sigma_e^2)$$

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

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

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

$$COV(a, c')=0; COV(a, e')=0; COV(c, e')=0, \text{ em que:}$$

A : matriz de parentesco genético aditivo.

I : matriz identidade de ordem apropriada.

σ_a^2 , σ_c^2 e σ_e^2 : variâncias genética aditiva, de ambiente comum da parcela e residual, respectivamente.

Sob este modelo, a herdabilidade (h^2) e a correlação devida ao ambiente comum da parcela (c^2) são, respectivamente:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2}; \quad c^2 = \frac{\sigma_c^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2}$$

A análise Bayesiana deste modelo linear mais geral (incluindo o efeito de parcela) é apresentada a seguir:

Distribuição conjunta a posteriori dos parâmetros do modelo

$$\begin{aligned} p(\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2 | y) &\propto p(\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2) p(y|\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2) \\ &= p(\beta) p(a|\sigma_a^2) p(c|\sigma_c^2) p(\sigma_a^2) p(\sigma_c^2) p(\sigma_e^2) p(y|\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2), \end{aligned}$$

Considerando a distribuição a priori dos componentes da variância como uma qui-quadrado escalonada invertida, tem-se que:

Distribuição conjunta a posteriori

$$p(\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2 | y) \propto \sigma_e^{2\left(\frac{n+\nu_e}{2}+1\right)} \exp\left[(y - X\beta - Za - Wc)'(y - X\beta - Za - Wc) + \nu_e S_e^2\right]$$

$$\begin{aligned} & \sigma_a^{2\left(\frac{q+\nu_a}{2}+1\right)} \exp\left[-\frac{(a' A^{-1} a + \nu_a S_a^2)}{2 \sigma_a^2}\right] \\ & \sigma_c^{2\left(\frac{p+\nu_c}{2}+1\right)} \exp\left[-\frac{(c' c + \nu_c S_c^2)}{2 \sigma_c^2}\right] \end{aligned}$$

Incorporando na matriz C (item 12.12) as submatrizes relacionando W com Z e X, obtém-se as distribuições condicionais a posteriori para cada parâmetro:

Distribuição condicional a posteriori de β_i

$$\beta_i | \beta_{-i}, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2, y \sim N(\hat{\beta}_i, (X'_{-i} X_{-i})^{-1} \sigma_e^2),$$

$$\text{em que: } \hat{\beta}_i = (X'_{-i} X_{-i})^{-1} X'_{-i} (y - X_{-i} \beta_{-i} - Za - Wc)$$

Distribuição condicional a posteriori de a_i

$$a_i | \beta, a_{-i}, c, \sigma_a^2, \sigma_c^2, \sigma_e^2, y \sim N(\hat{a}_i, (Z'_{-i} Z_{-i} + A_{-i,-i}^{-1} \alpha)^{-1} \sigma_e^2)$$

Distribuição condicional a posteriori de c_i

$$c_i | \beta, a, c_{-i}, \sigma_a^2, \sigma_c^2, \sigma_e^2, y \sim N(\hat{c}_i, (W'_{-i} W_{-i} + \alpha^*)^{-1} \sigma_e^2)$$

Distribuição condicional a posteriori de σ_a^2

$$\sigma_a^2 | \beta, a, c, \sigma_c^2, \sigma_e^2, y \sim \tilde{\nu}_a \tilde{S}_a^2 \chi_{\tilde{\nu}_a}^2$$

Distribuição condicional a posteriori de σ_c^2

$$\sigma_c^2 | \beta, a, c, \sigma_a^2, \sigma_e^2, y \sim \tilde{\nu}_c \tilde{S}_c^2 \chi_{\tilde{\nu}_c}^2$$

Distribuição condicional a posteriori de σ_e^2

$$\sigma_e^2 | \beta, a, c, \sigma_a^2, \sigma_c^2, y \sim \tilde{\nu}_e \tilde{S}_e^2 \chi_{\tilde{\nu}_e}^2 \text{ em que } \alpha^* = \frac{\sigma_e^2}{\sigma_c^2} \text{ e os demais parâmetros são definidos analogamente ao item 12.12.}$$

Além das distribuições consideradas para os efeitos aleatórios no modelo linear freqüentista e para a verossimilhança do vetor de observações, a abordagem Bayesiana requer atribuições para as distribuições a priori dos efeitos fixos e componentes de variância. Uma distribuição a priori não informativa ou uniforme foi atribuída a β , refletindo conhecimento a priori vago sobre os efeitos fixos, de forma que $p(\beta) \propto K$ (constante). Para os componentes de variância, distribuições χ^2 inversas foram consideradas como priori, conforme especificado no tópico anterior. Considerou-se adicionalmente $v_i = -2$ e $S_i^2 = 0$, de forma que a distribuição se tornasse uniforme, e, portanto, não informativa.

Para a inferência bayesiana sobre os parâmetros de interesse empregou-se a técnica da amostragem de Gibbs. O principal aspecto deste procedimento refere-se ao fato de as inferências basearem-se na distribuição marginal a posteriori dos parâmetros, sendo que a marginalização da distribuição conjunta a posteriori é obtida via o amostrador de Gibbs através de amostragens e atualizações das distribuições condicionais. Em resumo, a abordagem bayesiana baseia-se na construção da distribuição marginal a posteriori de um parâmetro de interesse tratando-o como uma variável aleatória e aplicando cálculo de probabilidades. Este procedimento implica problemas multidimensionais, uma vez que todos os outros parâmetros do modelo devem ser integrados (eliminados), fato que raramente é possível usando os métodos numéricos padrões. O procedimento iterativo da amostragem de Gibbs refere-se a uma técnica de integração estocástica que cria uma cadeia de Markov, que é uma distribuição (conjunta a posteriori) estacionária associada à distribuição a posteriori de interesse. Tomando-se amostras, iterativamente, das distribuições condicionais a posteriori, com contínua atualização, obtém-se a distribuição conjunta a posteriori em equilíbrio e, após um número de iterações suficientemente grande, a última amostra desta seqüência e qualquer amostra subsequente é uma amostra da distribuição conjunta a posteriori. Este resultado implica que cada coordenada do vetor de amostras retiradas, $\theta^n = [\beta^n \alpha^n c^n \sigma_a^{2(n)} \sigma_c^{2(n)} \sigma_e^{2(n)}]$, é uma amostra da distribuição marginal a posterioripropriada.

Para a análise bayesiana empregou-se o software MTGSAM (Van Tassell & Van Vleck, 1995) e para a análise clássica empregou-se o software MTDFREML (Boldman et al., 1995). Foram avaliados vários tamanhos de cadeia de Markov e intervalos entre amostragens.

Para análise de convergência do processo de estimação foram consideradas cadeias de tamanhos 10.000, 15.000 e 20.000 ciclos, intervalos de 50 e 100 entre amostras dos parâmetros de interesse e período de descarte de 2.000 ciclos. Alguns resultados referentes aos parâmetros genéticos são apresentados na Tabela 12.1.

Tabela 12.1. Resultados comparativos entre algumas estratégias de amostragem de Gibbs empregadas, para estimação (médias) dos parâmetros genéticos herdabilidade (h^2) e correlação de ambiente comum (c^2).

Tamanho da cadeia	Intervalo entre amostras	h^2	c^2
10.000	100	0,34	0,032
20.000	100	0,34	0,032
15.000	50	0,38	0,029
20.000	50	0,36	0,030

Os resultados apresentados na Tabela 12.1 juntamente com outros não apresentados permitiram concluir que a convergência do processo de amostragem ocorreu com intervalos de 100 entre amostras e com tamanho de cadeia de 10.000 ciclos (visto que a cadeia de 10.000 ciclos encontra-se dentro da cadeia de 20.000 ciclos e que ambas conduziram a resultados idênticos).

Outra forma de análise de convergência refere-se à estimativa do erro de Monte Carlo, que é uma estatística associada ao erro de estimativa de determinado parâmetro devido ao número de amostras utilizadas na cadeia de Gibbs, sendo que este erro é inversamente proporcional ao tamanho da cadeia (Van Tassel & Van Vleck, 1996). Este erro pode ser calculado pela variância dos parâmetros amostrados sucessivamente a cada intervalo dividida pelo número de amostras salvas, sendo que a raiz quadrada deste erro fornece uma aproximação para o desvio padrão do erro associado ao comprimento da cadeia. Considerando a cadeia de tamanho 20.000 e intervalo de 100 entre amostras, o desvio padrão do erro associado ao tamanho da cadeia foi de 0,0004 para a herdabilidade (Tabela 12.2), resultado que revela que os 18.000 ciclos, após o período de descarte, foram suficientes para obtenção de precisas estimativas das médias a posteriori. Dessa forma, optou-se por apresentar os resultados referentes ao tamanho de cadeia de 20.000 ciclos e com amostras salvas a cada 100 ciclos. O tempo de processamento para esta análise foi de cerca de 10 horas em um computador Pentium 100 Mhz, com 16 MB de memória.

Na Tabela 12.2 são apresentadas as estimativas de parâmetros genéticos pelos procedimentos REML e GS.

Tabela 12.2. Estimativas descritivas das distribuições marginais a posteriori da herdabilidade (h^2), correlação devida ao ambiente comum (c^2) e variância fenotípica (σ_F^2) obtidas pelo procedimento da amostragem de Gibbs (GS), bem como estimativas destes parâmetros obtidas pelo método da máxima verossimilhança restrita (REML). Caráter diâmetro em *Pinus caribaea* var. *hondurensis*.

Parâmetro	GS				REML	
	Média \pm s*	Moda	Mediana	S**	Intervalo de Confiança (95%)	Estimativa
h^2	0,34 \pm 0,005	0,34	0,34	0,0004	0,33 – 0,35	0,32 \pm 0,06
c^2	0,032 \pm 0,003	0,032	0,032	0,0002	0,026 – 0,038	0,031 \pm 0,01
σ_F^2	8,2441 \pm 0,02	-	-	-	-	8,0624

s* = desvio padrão; s** = desvio padrão Monte Carlo.

Os resultados pontuais obtidos pelas abordagens freqüentista e bayesiana foram similares (Tabela 12.2), conforme esperado. Uma vez que foram utilizadas distribuições a priori não informativas na análise bayesiana, as modas das distribuições marginais a posteriori dos parâmetros genéticos foram similares às correspondentes estimativas REML. Do ponto de vista bayesiano, as estimativas obtidas por REML correspondem às modas das distribuições conjuntas a posteriori dos componentes de variância, dada a utilização de prioris uniformes para os efeitos fixos e componentes de variância. A técnica GS possibilitou estimativas de componentes de variância pelo método VEIL (ou da verossimilhança integrada) apresentado por Gianola & Foulley (1990).

A grande vantagem da análise bayesiana, neste caso, refere-se à obtenção dos desvios padrões e intervalos de confiança exatos para os parâmetros genéticos e valores genéticos preditos. Verificou-se que a técnica GS propiciou estimativas muito mais precisas que aquelas obtidas por REML. Estimativas com precisões similares empregando o procedimento GS foram obtidas por Varona et al. (1994), para caracteres quantitativos em gado de leite e por Soria et al. (1998) em *Eucalyptus globulus*. Na Tabela 12.3 são apresentados os valores genéticos preditos dos melhores genitores, obtidos pelos procedimentos GS e BLUP, bem como os ganhos genéticos associados.

Tabela 12.3. Estimativas descritivas das distribuições marginais a posteriori dos valores genéticos preditos dos 10 melhores genitores obtidos pelo procedimento da amostragem de Gibbs (GS), bem como valores genéticos preditos pelo procedimento BLUP. Caráter diâmetro em *Pinus caribaea* var. *hondurensis*.

Genitor	Média ± Desvio Padrão	Ordem (GS)	BLUP ± SEP*	Acurácia (BLUP)	Ordem (BLUP)
130	2,70 ± 0,03	1	2,61 ± 0,80	0,86	1
49	2,23 ± 0,07	2	2,15 ± 0,94	0,81	3
104	2,19 ± 0,06	3	2,26 ± 0,86	0,84	2
73	2,07 ± 0,07	4	1,96 ± 0,99	0,78	5
114	2,03 ± 0,09	5	1,96 ± 1,06	0,75	6
56	1,99 ± 0,07	6	1,89 ± 0,96	0,80	7
100	1,97 ± 0,03	7	1,97 ± 0,78	0,87	4
92	1,90 ± 0,03	8	1,82 ± 0,79	0,87	8
76	1,80 ± 0,05	9	1,81 ± 0,96	0,80	9
125	1,79 ± 0,03	10	1,69 ± 0,65	0,91	10
		GS	BLUP		
Ganho genético (cm)	2,066 ± 0,04		(1,99 – 2,14)**	2,011 ± 0,28	(1,46 – 2,56)***
Ganho genético (%)	13,93 ± 0,27		(13,40 – 14,46)**	13,56 ± 1,88	(9,88 – 17,24)***

* desvio padrão do erro de predição.

** intervalo com 95% de confiança (probabilidade).

*** cálculo considerando uma acurácia média de 0,88.

As diferenças observadas nos valores genéticos preditos pelos procedimentos GS e BLUP/REML foram pequenas, conduzindo a uma ligeira alteração na ordem dos melhores pelos dois procedimentos (Tabela 12.3). Do ponto de vista bayesiano, uma vez que nenhuma distinção existe entre efeitos fixos e aleatórios, as soluções das equações de modelo misto do BLUP correspondem às médias das distribuições marginais a posteriori dos parâmetros de locação (efeitos fixos e aleatórios), dados os componentes de variância, ou parâmetros de dispersão. As condições para que haja esta correspondência entre BLUP e GS para os parâmetros de locação são: atribuição de prioris uniformes (não informativas) para os efeitos fixos, prioris normais para os efeitos aleatórios e verossimilhança normal para o vetor de observações. Estas premissas foram atendidas neste trabalho, fato que explica os resultados obtidos.

Segundo Wang et al. (1994), uma deficiência do procedimento BLUP/REML é que os erros de estimação dos componentes de dispersão não são considerados por ocasião da predição de valores genéticos. A abordagem bayesiana e a teoria de probabilidade a ela associada impõem que as inferências devem ser baseadas nas distribuições marginais a posteriori dos parâmetros de interesse, de forma que toda a incerteza sobre os parâmetros é considerada completamente. Isto é conseguido a partir da distribuição conjunta a poste-

riori de todos os parâmetros. Desta distribuição, a distribuição marginal a posteriori do valor genético de um indivíduo é obtida por sucessivas integrações (eliminações) de todos os parâmetros de *nuisance* do modelo, tais como efeitos fixos, outros efeitos aleatórios (ambiente comum) e componentes de variância e parâmetros associados (h^2 e c^2 no caso).

O aspecto ressaltado por Wang et al. (1994) é muito relevante considerando, principalmente, que para que as propriedades desejáveis do BLUP sejam asseguradas é necessário o conhecimento exato dos componentes de variância. Dessa forma, a análise Bayesiana, que permite estimativas quase exatas dos componentes de variância, tende a conduzir também a uma maior aproximação entre ganhos genéticos preditos e realizados com seleção. Este último aspecto pode, em parte, ser visualizado através dos desvios padrões dos ganhos genéticos apresentados na Tabela 12.3, sendo que o procedimento GS propiciou uma inferência muito mais segura (menor desvio padrão ou risco) sobre o ganho genético a ser capitalizado com a seleção dos 10 melhores genitores.

Os parâmetros genéticos iniciais utilizados foram obtidos pelo procedimento REML e as distribuições a priori foram tomadas como não informativas. Se as distribuições a priori para estes parâmetros fossem tomadas como informativas, menores desvios padrões para os parâmetros seriam obtidos. Quanto às estimativas pontuais (médias a posteriori), tem sido verificado que os valores obtidos são similares quando se consideram prioris muito informativas e não informativas (Van Tassell & Van Vleck, 1996).

As seguintes conclusões podem ser relatadas: (i) a análise Bayesiana propicia resultados adicionais àqueles obtidos pela abordagem freqüentista, destacando-se os intervalos de confiança bayesianos para as estimativas de parâmetros genéticos, valores genéticos e ganhos genéticos; (ii) as estimativas dos parâmetros genéticos, valores genéticos e ganhos genéticos pelo procedimento da amostragem de Gibbs podem ser mais precisas do que pelo procedimento REML/BLUP; (iii) a análise Bayesiana é uma técnica elegante e flexível que permite a simultânea estimação dos parâmetros genéticos, efeitos “fixos” e valores genéticos de maneira precisa, mesmo para amostras de tamanho finito. Entretanto, para obtenção de uma precisão adequada na inferência é necessário que o pesquisador tenha uma quantidade suficiente de informação experimental.

É importante relatar que o uso da abordagem Bayesiana em genética quantitativa não se restringe aos modelos lineares Gaussianos, mas, engloba também os modelos lineares generalizados (Tempelman, 1998), os modelos não lineares e os modelos robustos (Gianola, 2000). De maneira geral, além da riqueza de informações propiciada, a análise Bayesiana oferece maior flexibilidade quanto aos modelos de análise, visto que permite o ajuste de modelos de grande complexidade.

Capítulo 13

**Estatística Experimental:
Comparação entre Tratamentos
de Efeitos Fixos**

No contexto do melhoramento genético, os efeitos de tratamentos (cultivares, variedades, linhagens, progênies, populações, híbridos, clones) podem ser considerados fixos ou aleatórios. Em estatística, quando a variável resposta (y) é contínua (quantitativa) e os efeitos de tratamentos são considerados fixos, têm-se os modelos de análise de variância e de regressão, os quais são contemplados na maioria dos livros das várias áreas de aplicação da estatística: biometria, econometria, geoestatística e tecnometria (estatística aplicada à engenharia, física e química).

No caso da análise de variância (em que a variável indicadora (x) dos tratamentos é qualitativa), os modelos podem incluir efeitos fixos e aleatórios (por exemplo, tratamentos de efeitos fixos e blocos de efeitos aleatórios, no delineamento em blocos ao acaso) mas as análises enfatizam os efeitos fixos, e os efeitos aleatórios são vistos como fatores de *nuisance*, não havendo interesse em testá-los. O efeito de tratamento não é uma variável casual, mas sim um tratamento introduzido sob controle, de forma que uma função do somatório do quadrado dos efeitos dos tratamentos não pode ser considerada como uma variância. O interesse principal da análise, no caso, é estimar as magnitudes dos efeitos de tratamentos e também das diferenças entre tratamentos.

Nos modelos de regressão, a variável indicadora dos tratamentos é quantitativa, ou seja, os tratamentos referem-se a quantidades, dosagens ou níveis de fatores quantitativos (neste caso, o uso da análise de variância é reprovável) e o interesse reside no estudo do comportamento da variável resposta em face do incremento das dosagens. Para as variáveis categóricas, tem-se os modelos de regressão logística e de análise de tabelas de contingência apresentados na Tabela 13.1.

Tabela 13.1. Classificação das variáveis resposta (y) e indicadora ou preditora (x) e seus modelos de análise associados.

Variável y	Variável x	Modelo de análise
Contínua*	Categórica+	Análise de variância
Contínua*	Contínua*	Regressão
Categórica+	Contínua*	Regressão logística
Categórica+	Categórica+	Análise de tabelas de contingência

* Contínua e Quantitativa; + Categórica (Discreta) e Qualitativa.

A situação em que os tratamentos são assumidos como efeitos aleatórios caracteriza os modelos de componentes de variância. Não há interesse na magnitude dos efeitos de tratamentos e seus contrastes, mas sim na magnitude da variância entre tratamentos (σ_t^2) e seu valor em relação à variação residual (σ^2), em geral, expressa por $\sigma_t^2 / (\sigma_t^2 + \sigma^2)$, fato que torna relevante a estimativa de componentes de variância. Se os modelos incluem efeitos fixos, além do efeito aleatório de tratamento, estes são agora considerados de *nuisance*.

Assim, a rigor, estimar médias de tratamentos e realizar comparações de médias, quando os efeitos dos tratamentos são aleatórios não é teoricamente correto. Tanto na estimação dos efeitos fixos de tratamentos quanto na estimação de componentes de variância, os parâmetros de interesse referem-se a constantes (os parâmetros de locação e as variâncias paramétricas associadas às distribuições das variáveis aleatórias do modelo, respectivamente) que precisam ser estimadas. No caso em que os efeitos de tratamentos são aleatórios, eles podem ser preditos a partir dos dados observados e dos componentes de variância, caracterizando a predição de variáveis aleatórias, conforme abordado na maior parte deste documento.

A abordagem apresentada a seguir, refere-se aos modelos de análise de variância com tratamentos de efeitos fixos. O conteúdo apresentado refere-se à experimentação em geral e não apenas a tratamentos fixos associados à área de melhoramento genético, mas também a tratamentos e experimentos fitotécnicos, silviculturais, biológicos, químicos, físicos, etc. Considerações a respeito dos modelos de regressão logística são feitas no Capítulo 8 e sobre tabelas de contingência, no tópico 13.8.

13.1. Análise da Variância Paramétrica e Comparações Múltiplas

Considerando um modelo mais simples, como aquele associado à avaliação de tratamentos em um experimento no delineamento inteiramente casualizado, tem-se que:

$$Y_{ij} = \mu + t_i + e_{ij}$$

Y_{ij} : valor observado do i-ésimo tratamento na repetição j.

μ : efeito da média geral, fixo, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$.

t_i : efeito do tratamento i, fixo, $E(t_i) = t_i$ e $E(t_i^2) = t_i^2$.

e_{ij} : erro aleatório, $E(e_{ij}) = 0$ e $E(e_{ij}^2) = \sigma^2$, ou seja,

$e_{ij} \sim N(0, \sigma^2)$, de forma que os erros apresentam distribuição normal, são independentes e identicamente distribuídos com variância comum σ^2 .

Na especificação do modelo, verifica-se:

- (i) O modelo é aditivo nos parâmetros, ou seja, os efeitos da média geral, de tratamentos e do erro são aditivos e não correlacionados.
- (ii) Os erros apresentam distribuição normal.
- (iii) Os erros são independentes, ou seja, não apresentam correlação serial.
- (iv) Os erros são homocedásticos, ou seja, apresentam homogeneidade de variâncias.

Tal modelo associa-se a uma análise de variância do tipo:

FV	GL	QM	E(QM)	F
Tratamentos	t-1	Q ₁	$\sigma^2 + \frac{r}{t-1} \sum_{i=1}^t t_i^2$	Q ₁ / Q ₂
Erro	(r-1) t	Q ₂	σ^2	
Total	rt-1			

Considerando que uma razão entre variâncias tem distribuição F de Snedecor, o teste F, em princípio, destina-se à comparação entre variâncias. No entanto, presta-se também à comparação de $\sigma^2 + \frac{r}{t-1} \sum_{i=1}^t t_i^2$ com σ^2 , ou seja, no teste da igualdade de Q₁ com Q₂. Neste caso, se for comprovado que Q₁ = Q₂, tem-se $\sum_{i=1}^t t_i^2$ estatisticamente igual a zero, fato que permite concluir que os dados podem ser tratados como uma única amostra de rt itens provenientes de $Y \sim N(\mu, \sigma^2)$ ou, em termos práticos, que não existem diferenças reais entre os tratamentos.

Formalmente, é necessário formular as hipóteses estatísticas, as quais podem ser denominadas: (i) H₀: **hipótese de nulidade** ou de igualdade de tratamentos; (ii) H_a: **hipótese alternativa**, hipótese de pesquisa ou do pesquisador ou hipótese da diferença entre tratamentos. A hipótese H₀ pode ser verdadeira ou falsa, e a aceitação ou rejeição desta hipótese por parte do pesquisador está sempre associada a um grau de incerteza na tomada da decisão sobre tal hipótese estatística.

Associada a esta decisão, o pesquisador pode cometer os erros tipo I e tipo II, conforme ilustrado a seguir.

Conclusão do Teste Estatístico		Situação real na população	
		H ₀ verdadeira	H ₀ falsa
Teste Estatístico	Rejetar H ₀	Erro Tipo I → α	Decisão Correta → (1-β)
	Aceitar H ₀	Decisão Correta → (1-α)	Erro Tipo II → β

Assim, o pesquisador incorre no erro tipo I, quando rejeita uma hipótese H₀ que é verdadeira e incorre no erro tipo II, quando aceita uma hipótese H₀ que é falsa. A probabilidade de cometer um erro tipo I é designada por α e o maior valor de α para H₀ verdadeira é denominado **nível de significância** de um teste estatístico, ou seja, a significância de um teste é a probabilidade máxima que se admite correr o risco de cometer um erro tipo I. A probabilidade (1-α) é denomi-

nada **grau de confiança**. A probabilidade de cometer um erro tipo II é designada por β . A probabilidade $(1-\beta)$ é denominada **poder** de um teste e refere-se à probabilidade de se rejeitar H_0 , quando H_0 é falsa, ou seja, à capacidade de detectar as diferenças reais que existem entre tratamentos.

Na prática, o pesquisador tem controle apenas sobre o tipo I já que pode escolher, nominalmente, o nível de significância α . Quanto ao erro tipo II não há como controlá-lo. No caso do teste F, pode ser estabelecida a hipótese:

$$H_0 = \sum_{i=1}^r t_i = 0, \text{ ou seja, os efeitos de tratamentos não diferem entre si.}$$

Caso esta hipótese seja rejeitada a um nível de significância α , conclui-se que existe pelo menos um contraste (diferença) significativo entre duas médias, pelo teste t de Student. Sendo a média de cada tratamento estimada por

$$\bar{Y}_{i\cdot} = \sum_{j=1}^r Y_{ij} / r \text{ e a variância da média de tratamento, por } \hat{Var}(\bar{Y}_{i\cdot}) = \hat{\sigma}^2 / r$$

(em que $\hat{\sigma}^2 = Q_2$), pode-se comparar todas as médias de tratamentos mediante procedimentos de comparações múltiplas, os quais fazem uso de $\bar{Y}_{i\cdot}$ e de $(\hat{\sigma}^2 / r)$.

Existem vários procedimentos para comparações múltiplas (Steel & Torrie, 1980; Neter et al., 1996). Dentre as características desejáveis destes procedimentos, sobressaem a baixa probabilidade de erro tipo I e o alto poder dos testes, ambos conceitos associados à rejeição da hipótese H_0 . O controle do erro tipo I pode ser realizado por comparação (a probabilidade de erro tipo I é dada pela razão entre o número de erros tipo I, em comparações duas a duas e o número total de comparações envolvendo todas as combinações duas a duas entre tratamentos em todos os experimentos) ou por experimento (a probabilidade de erro tipo I é dada pela razão entre o número de experimentos com pelo menos um erro tipo I e o número total de experimentos). Segundo Perecin & Barbosa (1988), a questão prática é que um teste com taxa de erro por comparação torna-se muito frouxo ao ser aplicado em todo o experimento, enquanto que um teste com taxa de erro por experimento torna-se conservador ou rigoroso ao ser olhado por comparação. Uma explicação simples e acessível é dada por Vieira (1999).

Os principais testes de comparações múltiplas são: t (ou LSD de Fisher ou forma F protegida de Fisher, quando aplicado somente quando o teste F é significativo), Tukey (ou HSD), Duncan, Scheffé, Dunnet, Newman-Keuls (ou Student-Newman-Keuls – SNK ou simplesmente Keuls), Bonferroni (ou LSDB) e Scott-Knott. Destes, o teste de Dunnet é adequado para comparação das médias dos tratamentos com as médias das testemunhas e os demais permitem comparações entre as médias dos vários tratamentos. É importante destacar que todos estes testes foram derivados na suposição de tratamentos de efeitos fixos e,

portanto, não se justifica a sua aplicação quando os efeitos de tratamentos são considerados aleatórios.

Os testes t não protegido e Duncan controlam as taxas de erro tipo I por comparação, ao passo que os testes de Tukey e Scheffé controlam as taxas de erro tipo I, por experimento (Chew, 1986). Assim, ao serem olhados por comparação, os testes menos rigorosos são o t e o Duncan e os mais rigorosos são o Tukey e o Scheffé. O teste de Newman-Keuls controla a taxa de erro por comparação mas não controla totalmente a taxa de erro por experimento, apresentando, portanto, um rigor intermediário. Assim, dos menos rigorosos para os mais rigorosos têm-se: t, Duncan, Newman-Keuls, Tukey, Scheffé. O rigor, no caso, refere-se à maior dificuldade em detectar diferenças significativas entre tratamentos. Dessa forma, a um maior rigor está associado também um menor poder. Como o nível de significância e o poder de um teste crescem juntos, uma alternativa para aumentar o poder de testes muito rigorosos é aumentar o nível de significância, por exemplo, adotando $\alpha = 10\%$.

O teste de Bonferroni refere-se a uma modificação do teste t não protegido, visando conservar a taxa de erro tipo I, por experimento. Neste caso, o rigor em relação ao teste t é aumentado. O teste de Scott-Knott controla bem a taxa de erro tipo I, por experimento e por comparação, conservando-a em torno ou abaixo do nível nominal de significância (Silva et al., 1999).

Outro teste cujo uso é freqüente é o t bayesiano ou teste de Waller-Duncan. Este teste não envolve o conceito usual de taxas de erro e também utiliza o valor calculado da estatística F no cálculo do valor crítico de t^* . Se o valor de F calculado é grande (indicando tratamentos heterogêneos), o valor crítico é reduzido, tornando o teste mais poderoso. Se o valor calculado de F é pequeno, o valor crítico de t^* será aumentado, tornando o teste menos poderoso, ou seja, mais rigoroso (Chew, 1986).

Uma avaliação (via simulação) de vários testes quanto às taxas de erro tipo I e poder foi realizada por Perecin & Barbosa (1988) e revelou os seguintes resultados:

- Os testes t e Duncan não apresentam taxas favoráveis de erro tipo I, ou seja, não conservam adequadamente os níveis nominais de significância.
- O teste de Tukey possui um poder muito reduzido, especialmente quando há grande número de tratamentos.
- O teste t bayesiano concilia de certa forma as características de alto poder e baixas taxas de erro tipo I, entretanto, as taxas de erro tipo I não podem ser previstas com exatidão e dependem do número de tratamentos e de suas magnitudes, requerendo, portanto, aplicação cuidadosa.
- O teste de Newman-Keuls possui poder muito superior e taxas de erro tipo I similares ao teste de Tukey.

Assim, os testes t, Duncan e Tukey, amplamente utilizados no Brasil, não são os mais recomendados e só devem ser usados com ressalvas. O teste de Newman-Keuls, pouco utilizado no Brasil, é altamente recomendado tendo em vista as taxas favoráveis de erro tipo I, o relativo alto poder e o rigor intermediário. Dessa forma, pode ser utilizado sem maiores cuidados. Em realidade, tal teste vem sendo amplamente utilizado (em detrimento dos demais) pelos franceses e também nos trabalhos de melhoramento de plantas perenes em países da África.

A título de exemplo, será considerada a aplicação do teste de Newman-Keuls. Este teste usa a amplitude estudentizada ou padronizada (q) (assim como o fazem os testes Tukey e Duncan) e é decorrente dos trabalhos de Newman (1939) e Keuls (1952).

Todos os testes baseiam-se na especificação de uma diferença mínima significativa (DMS ou LSD) entre as médias de tratamentos, a qual está associada a um nível de significância α . Esta DMS, em conjunto com uma **estatística de teste** dada pelo contraste de médias de dois tratamentos, gera uma **regra de decisão**. A DMS permite testar a hipótese $H_0 : (\mu + t_i) - (\mu + t_j) = 0$, ou seja, a hipótese da nulidade ou da igualdade das médias de tratamentos. Se o contraste de médias for superior à DMS, rejeita-se a hipótese H_0 e os tratamentos são diferentes estatisticamente. Esta é a **regra de decisão**.

Para o teste de Newman-Keuls esta DMS é dada por:

$$DMS = q_{(m,\eta,\alpha)} [\hat{\sigma}/(r)^{1/2}] = q_{(m,\eta,\alpha)} [Q_2 / r]^{1/2}, \text{ em que:}$$

$q_{(m,\eta,\alpha)}$ valor da amplitude padronizada (que pode ser obtida de uma tabela para o teste de Tukey) associada ao número m de médias ordenadas abrangidas pela comparação, ao número η de graus de liberdade do resíduo da análise de variância e nível de significância α (1%, 5% ou 10%).

$[Q_2 / r]^{1/2} = \hat{\sigma}/(r)^{1/2}$: erro padrão da média de tratamentos, função do quadrado médio do resíduo (Q_2) da análise de variância.

Esta expressão é similar às expressões da DMS empregadas pelos testes de Duncan e de Tukey. Difere da expressão do teste de Duncan devido ao fato de o nível de significância α não ser constante no teste de Duncan (este também é um ponto desfavorável deste teste) e ser constante para o teste de Newman-Keuls. Difere do teste de Tukey em razão de o valor m ser constante no teste de Tukey e variável no teste de Newman-Keuls, de acordo com o número de médias abrangidas pela comparação ou contraste.

Considerando os dados apresentados por Steel & Torrie (1980), têm-se as seguintes médias para 6 tratamentos: $T_1 = 13,3$; $T_2 = 14,6$; $T_3 = 18,7$; $T_4 = 19,9$; $T_5 = 24,0$ e $T_6 = 28,8$. Sendo o quadrado médio do resíduo $Q_2 = 11,79$, $r = 5$ repetições e $\eta = 24$, têm-se os seguintes valores de $q_{(m, 24, 0,05)}$ e de DMS para um nível de significância $\alpha = 5\%$:

m	$q_{(m, 24, 0,05)}$	DMS
2	2,92	4,5
3	3,53	5,4
4	3,90	6,0
5	4,17	6,4
6	4,37	6,7

Comparando-se inicialmente T_1 com T_6 , tem-se $T_6 - T_1 = 15,5$ e $DMS_6 = 6,7$ (esta é a DMS única que seria usada no teste de Tukey), de forma que T_6 é significativamente diferente de T_1 . As próximas comparações a serem realizadas são, então, T_1 com T_5 e T_2 com T_6 . Sendo $T_5 - T_1 = 10,7$ e $T_6 - T_2 = 14,2$ e a $DMS_5 = 6,4$, conclui-se que $T_1 \neq T_5$ e $T_2 \neq T_6$. As próximas comparações a serem feitas são T_4 com T_1 e T_6 com T_3 . Sendo $T_4 - T_1 = 6,5$ e $T_6 - T_3 = 10,1$ e a $DMS_4 = 6,0$, conclui-se que $T_4 \neq T_1$ e $T_6 \neq T_3$. Em seguida, faz-se as comparações de T_3 com T_1 e de T_6 com T_4 , donde concluiu-se que $T_1 = T_3$ e $T_6 \neq T_4$, uma vez que $T_3 - T_1 = 5,4$, $T_6 - T_4 = 8,9$ e $DMS_3 = 5,4$.

Neste ponto concluiu-se que $T_1 = T_2$ e $T_2 = T_3$, não sendo necessário testar os referidos contrastes. Entretanto, falta contrastar ainda T_4 com T_5 e T_5 com T_6 , visto que concluiu-se que $T_6 \neq T_4$. Sendo $T_5 - T_4 = 4,1$, $T_6 - T_5 = 4,8$ e $DMS_2 = 4,5$, concluiu-se que $T_5 = T_4$ e $T_5 \neq T_6$. Em resumo, tem-se:

T_1	T_2	T_3	T_4	T_5	T_6
13,3	14,6	18,7	19,9	24,0	28,0

Verifica-se que tal teste conduziu a ambigüidade nos resultados, conforme mostrado pelas sobreposições das barras que unem os tratamentos estatisticamente iguais. Isto ocorre também para os demais testes, exceto para o de Scott & Knott (1974). Tal teste é baseado em uma análise de agrupamento, a qual conduz a grupos mutuamente exclusivos de tratamentos e, portanto, elimina o problema de ambigüidade. Além desta propriedade desejável, o teste de Scott-Knott também é favorável em termos das taxas de erro tipo I e apresenta poder superior aos demais testes (Silva et al., 1999), sendo por isto fortemente recomendado. Ainda segundo Silva et al. (1999), este teste é conservador ou rigoroso mas não tanto quanto os testes de Scheffé, Tukey & Newman-Keuls.

Todas as considerações realizadas neste tópico podem ser extrapoladas também para o delineamento em blocos e para a comparação de médias em outros arranjos (fatorial, parcela subdividida, etc) experimentais.

13.2. Alternativas aos Procedimentos de Comparações Múltiplas

Conforme Chew (1986), os procedimentos de comparações múltiplas são muito mais utilizados nos Estados Unidos do que na Inglaterra. Para determinados estatísticos de renome como J. A. Nelder (ex-chefe do Departamento de Estatística da Estação Experimental de Rothamsted, Inglaterra, após R.A. Fisher e F. Yates que foram 1º e 2º chefes), os métodos de comparações múltiplas não têm importância na interpretação dos dados. Preece (1982) também comenta que não utiliza procedimentos de comparações múltiplas por não ver vantagens em utilizá-los na interpretação dos resultados de experimentos agrícolas.

O sentimento geral desses estatísticos refere-se ao fato de os experimentos serem realizados, principalmente, com objetivos de avaliar a magnitude dos efeitos de tratamentos e não apenas testar se esses efeitos diferem. Assim, torna-se, por vezes, mais importante verificar se os tratamentos diferem por uma diferença d especificada pelo experimentador, de acordo com a importância prática desta diferença.

Segundo Chew (1986), para propósitos de publicação de artigos científicos, basta apresentar uma tabela contendo as médias dos tratamentos seguidas por uma diferença mínima significativa (1,96 vezes o desvio padrão por exemplo) ou pelos desvios padrões destas médias. Estas informações são suficientes para o leitor escolher entre testar uma hipótese ou realizar uma estimativa intervalar.

Outras opções, muito úteis aos melhoristas são: (i) a quantificação da probabilidade da maior média amostral de tratamento referir-se, de fato, ao melhor tratamento, em termos de valor paramétrico; (ii) a determinação do número de tratamentos (genótipos no caso) a ser selecionado, de forma que o grupo selecionado conte com o melhor tratamento (genótipo).

Na situação (i), a referida probabilidade será alta se o número de repetições for suficientemente grande. Dada uma probabilidade P e uma diferença d (de importância prática) entre os dois melhores tratamentos, Dudewicz (1976) fornece tabelas para obtenção do tamanho necessário da amostra para atingir uma probabilidade P desejável. Esta tabela pode ser utilizada de forma reversa (dado o número de repetições utilizado na experimentação) para inferir sobre a probabilidade de o tratamento com maior média amostral ser de fato o melhor tratamento, para uma diferença d especificada.

A opção (ii) pode ser estruturada considerando uma probabilidade P de que o melhor tratamento esteja incluído em um grupo de tratamentos selecionados. Neste caso, a regra para a inclusão de um determinado tratamento dentre os selecionados é dada pela comparação da média amostral do tratamento com a estatística ($M_{max} - LSD$), em que M_{max} refere-se à maior de todas as médias do experimento e LSD, à DMS do teste de Dunnett. Assim, se $(M_{max} - LSD) = 15$ para $P = 95\%$ e as médias dos 4 supostamente melhores tratamentos são 18,

17, 16 e 14, torna-se necessário selecionar os 3 melhores tratamentos para ter 95% de confiança de estar selecionando o melhor tratamento.

Esta mesma abordagem pode ser aplicada na seleção dentro de progêneres, visando ter uma alta confiança de que o melhor indivíduo da progênie está sendo selecionado. Neste caso, no cômputo da estatística de teste, devem ser utilizados os valores genéticos preditos dos indivíduos e suas variâncias do erro de predição.

É interessante notar que o que se aplica aqui é a filosofia do teste de Dunnett, e que a inclusão do tratamento ou indivíduo com média amostral ou valor genético predito M_i é ditada pela regra $(M_{max} - M_i) < DMS$. No caso, M_{max} funciona como a média do controle ou testemunha e a DMS é dada por $DMS = d_{(\eta, \alpha)} [\hat{\sigma}/(r)^{1/2}] = d_{(\eta, \alpha)} [Q_2/r]^{1/2}$, em que d é o valor tabelado associado ao teste de Dunnett.

13.3. Pressupostos da Análise de Variância e suas Verificações

Na especificação do modelo linear, realizada no tópico anterior, mostrou-se que são quatro os pressupostos da análise de variância: aditividade dos efeitos do modelo, normalidade dos erros, independência dos erros e homocedasticidade ou homogeneidade de variância dos erros. Fica assim, explícita, a importância da análise de resíduos no contexto da análise de variância. Erros ou resíduos referem-se a discrepâncias entre os valores observados e preditos pelo modelo.

13.3.1. Aditividade

Este é um dos pressupostos mais importantes, pois, se violado, indica inadequação do modelo aos dados ou inadequação dos dados ao modelo. A aditividade pode ser verificada pelo método de Tukey (1949), o qual desdobra a soma de quadrados do erro em um grau de liberdade para a não aditividade e $(\eta - 1)$ graus de liberdade para o resíduo da não aditividade, onde η são os graus de liberdade do erro. O quadrado médio da não aditividade é testado contra o resíduo da não aditividade, pelo teste F.

Outro método relevante no estudo da aditividade é o da concorrência de Mandel (1961, 1971), especialmente indicado para estudos de aditividade em modelos relativos à interação genótipo x ambiente.

A detecção de não aditividade pode revelar a necessidade de transformação de dados, sendo indicada a transformação logarítmica para os casos de modelos multiplicativos.

13.3.2. Normalidade

A não normalidade dos dados e erros é restritiva apenas em situações muito drásticas, especialmente quando o teorema central do limite não se aplica,

ou seja, quando a distribuição da média dos dados não converge para a distribuição normal com o aumento do tamanho da amostra. Quando a distribuição dos erros é muito assimétrica (coeficiente de assimetria muito diferente de zero) e platicúrtica ou achatada (coeficiente de curtose muito negativo), a não normalidade afeta o nível nominal de significância, de forma que o nível escolhido pelo pesquisador não é assegurado na prática, usualmente sendo maior que o nominal, podendo ser constatadas diferenças significativas entre tratamentos que, em realidade, são iguais.

De maneira genérica, o teste F de Snedecor é robusto a pequenos desvios da normalidade dos erros, fornecendo resultados satisfatórios quando se verificam distribuições aproximadamente normais. Na presença de grandes desvios da normalidade e de dados com estrutura não linear, a aplicação direta dos modelos lineares sobre os dados observados torna-se imprópria. Neste caso, há duas opções: (i) a transformação dos dados de forma que se tornem adequados (ou se moldem) ao modelo linear; (ii) a moldagem dos modelos aos dados. A abordagem (ii) é preferida e caracteriza a técnica de modelos lineares generalizados (ver Capítulo 8). O uso desta técnica conduz, na pior das hipóteses, aos mesmos resultados que a abordagem linear.

Variáveis dicotômicas e categóricas, por vezes, não apresentam normalidade de erros. Estas variáveis são, provavelmente, uma aproximação da variável real de interesse, e, muitas vezes, as categorias surgem porque não é possível medir a variável real de interesse. Tomar as variáveis categóricas como normais é tanto mais apropriado 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 às variáveis.

Existem vários procedimentos para avaliar a normalidade de dados. Dentro eles, destacam-se os testes não paramétricos de Kolmogorov-Smirnov, de Shapiro-Wilks e de aderência do qui-quadrado (χ^2), que são os mais utilizados. O teste de aderência do χ^2 permite verificar ajustamento de uma distribuição de freqüências de uma amostra a uma distribuição teórica, no caso a distribuição normal padrão. A estatística χ^2 é dada por $\chi^2 = \sum_{i=1}^c (O_i - E_i)^2 / E_i$, em que O_i são

os valores observados na classe i da distribuição de freqüências, E_i refere-se ao valor esperado (sob H_0 , ou seja, sob a hipótese de normalidade) na classe i, com base na distribuição normal padrão, e c refere-se ao número de classes. O valor calculado do χ^2 é comparado com o valor tabelado de uma distribuição χ^2 com $(c - k - 1)$ graus de liberdade, adotando-se determinado nível de significância. K refere-se ao número de parâmetros estimados, no caso k = 2, tendo em vista que são estimados a média e o desvio padrão para o cômputo da variável padronizada z_i . Sendo o valor de χ^2 menor que o valor tabelado, aceita-se a hipótese de que os dados pertencem a uma distribuição normal.

O teste de Shapiro & Wilk (1965) pode ser aplicado segundo os seguintes passos, para o caso de amostras com tamanho $n < 50$:

- a) ordenamento crescente das observações $x_1 < x_2 \dots < x_n$.
- b) cálculo da variância $\hat{\sigma}^2$ das observações.
- c) obtenção de k : $k = n/2$ se n é par e $k = (n-1)/2$ se n é ímpar.
- d) obtenção de:

$$b = a_n(x_n - x_1) + a_{n-1}(x_{n-1} - x_2) + \dots + a_{n-k+1}(x_{n-k+1} - x_k),$$

em que os valores de a_{n-i+1} são tabelados para $n = 3, 4 \dots 50$.

- e) obtenção da estatística de teste $W = b^2 / \hat{\sigma}^2$.
- f) compara-se o valor calculado de W com o tabelado e sob a hipótese H_0 de normalidade, aceita-se a normalidade, se W calculado $< W$ tabelado.

O teste de Komogorov-Smirnov, a exemplo do χ^2 de aderência, também compara as distribuições dos dados observados com a dos dados esperados sob suposição de normalidade padrão. Este teste baseia-se na maior diferença D entre as freqüências acumuladas observadas e esperadas e tende a ser mais poderoso que o teste χ^2 , sobretudo no caso de pequenas amostras. A estatística do teste D é comparada com os valores críticos tabelados. Maiores detalhes sobre vários testes não paramétricos são apresentados por Siegel (1956).

Constatando-se grande desvio de normalidade e não se optando pelo uso da técnica de modelos lineares generalizados via análise de *deviance*, resta realizar a transformação dos dados. Persistindo o problema, devem ser empregadas as provas de livre distribuição, ou seja, os procedimentos não paramétricos. Os testes não paramétricos devem ser usados como último recurso, visto que os testes paramétricos têm melhores propriedades estatísticas, principalmente o poder.

13.3.3. Independência

A suposição de independência dos erros deve, em geral, ser obedecida no contexto da análise de variância. Isto porque os modelos de análise de variância assumem uma estrutura de covariância residual com variâncias ou correlações seriais nulas.

As correlações seriais entre observações (ou erros) surgem a partir da tomada de dados seqüenciais no tempo e/ou no espaço em um mesmo tratamento ou em diferentes tratamentos. São, então, gerados dados temporalmente e/ou espacialmente dependentes. As consequências dessas correlações na análise de variância são sérias e podem conduzir a erros nas inferências sobre as médias.

Apesar de existir um teste formal (teste de Durbin-Watson) para avaliar a ocorrência de correlação serial entre os erros, é importante destacar que tal correlação resulta da não aleatoriedade na obtenção da amostra de dados. Assim, a suposição de independência pode ser verificada simplesmente observando se

as unidades amostrais foram obtidas independentemente. Neste caso, é necessário saber exatamente como foram obtidos os dados, tendo em vista o planejamento e execução da amostragem e o objetivo da investigação. A casualização dos tratamentos nas unidades experimentais conduz à independência dos erros.

Em vários casos, são gerados dados temporalmente e/ou espacialmente dependentes como nos estudos das séries temporais, no estudo de variáveis regionalizadas em geoestatística, nos procedimentos de simulação estocástica (amostragem de Gibbs, por exemplo), nos procedimentos de reamostragem (Bootstrap e Jackknife) e em alguns casos de avaliação de medidas repetidas e dados longitudinais. Neste caso, devem ser utilizados modelos especiais que contemplam (por exemplo, via linhas e colunas auto-regressivas), em suas estruturas de variâncias, os erros correlacionados. Estes modelos são, via de regra, adotados nas análises de séries temporais e em geoestatística. Nos casos da simulação estocástica e dos processos de reamostragem, devem ser adotados procedimentos que quebrem a correlação serial entre as observações, como o descarte de amostras sucessivas e a tomada de amostras úteis a determinados intervalos (caso da amostragem de Gibbs) ou a utilização do método de Quenouille-Tukey para remover vícios devidos à autocorrelação (caso da reamostragem). Na análise de medidas repetidas, a presença de autocorrelação ou correlação serial (correlações aumentando com a diminuição da distância temporal entre amostras) pode indicar a necessidade do uso da análise de variância multivariada.

13.3.4. Homocedasticia

A homocedasticia ou homogeneidade de variância dos erros associados aos vários tratamentos é muito importante no contexto da comparação de médias, visto que os testes de comparação múltipla baseiam-se em diferenças mínimas significativas, dependentes de uma variação residual comum a todos os tratamentos. Assim, os erros contribuídos pelos vários tratamentos devem, todos, ser estimativas de uma variância populacional comum.

A violação de qualquer das outras suposições da análise de variância pode conduzir à heterogeneidade de variâncias dos erros. Segundo Scheffé (1959), a violação da suposição de homogeneidade de variâncias é grave quando a distribuição dos erros é leptocúrtica (curtose positiva) ou quando existe assimetria, e, no caso de distribuição leptocúrtica, o teste F tende a não rejeitar a hipótese da nulidade (H_0), quando ela é falsa.

Existem vários testes para inferências sobre a existência ou não de homogeneidade de variâncias, como o de Bartlett, o de Hartley e o de Levene. O teste de Bartley para comparação de variâncias é muito sensível à falta de normalidade dos erros, sendo por isto muito criticado. O teste de Hartley ou teste do F máximo tem como estatística de teste $F_{\max} = \hat{\sigma}_{\max}^2 / \hat{\sigma}_{\min}^2$ com t e (b - 1) graus

de liberdade, em que $\hat{\sigma}_{\max}^2$ e $\hat{\sigma}_{\min}^2$ referem-se à maior e à menor estimativas das variâncias residual (ou dentro de tratamentos) dos tratamentos ou amostras, t diz respeito ao número de tratamentos ou amostras e b, ao número de observações (ou blocos) por tratamento. O valor calculado do F_{\max} é comparado com o valor tabelado da distribuição de F_{\max} de Hartley. Como regra prática tem sido aceito que quando F_{\max} calculado é menor que 3, a heterogeneidade de variância não é problemática.

O teste de Levene consiste na análise de variância e teste F dos resíduos (valores observados em cada parcela menos a média de cada tratamento) dos vários tratamentos. Se o teste F não indicar diferenças significativas entre os resíduos dos tratamentos, aceita-se a homogeneidade de variância dos resíduos. Este teste tem grande aplicação (van Valen, 1978) e vem substituindo, vantajosamente, os demais.

No caso em que a heterogeneidade de variâncias é comprovada, têm-se as opções de padronização e de transformação dos dados. Segundo Steel & Torrie (1980), a heterogeneidade de variâncias pode ser classificada como regular e irregular. O tipo irregular caracteriza-se pela inexistência aparente de relação entre médias e variâncias. Por outro lado, a heterogeneidade regular surge de alguma forma de não normalidade, caracterizando-se por um relacionamento (correlação) entre médias e variâncias. Neste último caso, a transformação se aplica bem, desde que a distribuição dos dados seja conhecida.

Segundo Scheffé (1959), o balanceamento do experimento, ou seja, a adoção de números iguais de repetições por tratamento é um grande passo na proteção contra os efeitos da heterogeneidade de variâncias.

13.3.5. Transformações de dados

As transformações de dados podem, em casos específicos, resolver os problemas de não normalidade, heterogeneidade de variâncias e não aditividade. Transformações apropriadas podem gerar dados com distribuição aproximadamente normal e com independência entre médias e variâncias, resultando em variâncias homogêneas. Os principais tipos de transformação são a logarítmica, a raiz quadrada e a arco-seno ou angular.

A transformação logarítmica estabiliza a variância, na situação em que as variâncias são proporcionais ao quadrado das médias dos tratamentos. Em alguns casos, pode contribuir para a normalização dos dados e para a adequação do modelo aditivo linear.

A transformação raiz quadrada é indicada para estabilizar ou homogeneizar a variância quando existe correlação entre média e variância e a variável refere-se a uma contagem, com distribuição de Poisson (esta distribuição tem média igual a variância). Neste caso, a variável transformada pode ser considerada com distribuição normal.

A transformação arco-seno é aplicável a dados com distribuição binomial, expressos em frações ou porcentagens, ocorrendo estabilização da variância. Em geral, quando todos os dados equivalem a porcentagens apenas na faixa de 30 a 70, a transformação provavelmente não seja necessária.

Para dados discretos, em geral, recomenda-se verificar a existência ou não de correlação entre as médias de cada tratamento e suas variâncias. Se for constatada tal correlação, deve-se identificar a distribuição (Binomial ou Poisson) dos dados e aplicar a transformação recomendada.

13.4. Análise de Variância Não Paramétrica

Nas situações em que os pré-requisitos para a análise de variância paramétrica são seriamente violados e as transformações de dados não surtem efeito, deve-se lançar mão da inferência estatística não paramétrica. Neste caso, os testes não paramétricos são mais eficientes que os paramétricos.

Esta modalidade de inferência baseia-se em testes de livre distribuição, ou seja, testes cujos modelos não especificam a distribuição da população da qual os dados foram obtidos. Quando todas as suposições do modelo estatístico são satisfeitas, a prova paramétrica é mais poderosa que a não paramétrica. Entretanto, segundo o conceito de **poder-eficiência**, mesmo quando todas as suposições da prova paramétrica forem satisfeitas, a prova não paramétrica pode conduzir ao mesmo poder na rejeição de H_0 , por meio de conveniente aumento do tamanho da amostra. Assim, o conceito **poder-eficiência** pode ser utilizado na avaliação da eficiência comparativa entre os testes paramétricos e não paramétricos.

O conceito de **poder-eficiência** está associado ao aumento do tamanho da amostra para uma prova B, necessário para tornar esta prova tão poderosa quanto uma prova A. Assim, o **poder-eficiência** é dado por:

$$\text{poder-eficiência do teste } B = \frac{N_A}{N_B} \times 100,$$
 em que N_A e N_B são os tamanhos associados às

amostras para as provas A e B, respectivamente.

No caso, a prova A refere-se à paramétrica, aplicada a dados que satisfazem suas condições e B é a prova não paramétrica alternativa, ou seja, adequada aos mesmos objetivos e funções do teste. Se uma prova B exige uma amostra com $N_B = 30$ para ter o mesmo poder que uma prova A com $N_A = 27$, então o teste B tem **poder-eficiência** de 90%, significando que para igualar o poder do teste não paramétrico B com o teste paramétrico A é preciso uma amostra de 10 elementos para o teste B, para cada 9 elementos para o teste A.

13.4.1. Análise de variância de Kruskal-Wallis para o delineamento inteiramente casualizado

A prova de Kruskal-Wallis (Kruskal & Wallis, 1952) deve ser utilizada no caso de delineamentos em classificação simples ou única, ou seja, em situa-

ções de ensaios inteiramente casualizados. Neste caso, a análise de variância ou teste de Kruskal-Wallis é um substituto do teste F e testa a hipótese nula de que k amostras ou tratamentos independentes não diferem entre si. O método supõe que a variável analisada tenha uma inerente distribuição contínua e exige uma escala de mensuração ao menos ordinal.

Neste teste, cada uma das N observações é substituída por um posto ou ordem, e o método consiste dos seguintes passos:

- Ordenamento de todas as observações para os k tratamentos em uma única série de 1 a N, da menor para a maior.
- Determinação da soma de postos por coluna, ou seja, para cada um dos k tratamentos, sendo esta soma denominada R_j .
- Cômputo da estatística de Kruskal-Wallis, dada por:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1), \text{ em que } n_j \text{ é o número de observações no tratamento } j.$$

- Verificação da significância da estatística H, a qual tem distribuição χ^2 com $k - 1$ graus de liberdade, quando $k > 3$ ou os números de observações por tratamento não são muito pequenos, ou seja $n_j > 5$. Neste caso, rejeita-se H_0 se H for maior ou igual o valor tabelado de χ^2 . No caso de $k = 3$ e amostras pequenas ($n < 5$), outro procedimento (Siegel, 1956) para a verificação da significância deve ser utilizado.

Quando ocorrem empates entre dois ou mais escores, atribui-se a média dos postos a cada um deles. Na situação em que ocorre uma grande ($>$ que 25%) proporção de observações com valores empatados, a fórmula de H a ser utilizada equivale a:

$$H = \frac{\frac{12}{N(N+1)} \sum_{j=1}^k R_j - 3(N+1)}{1 - \frac{\sum T}{N^3 - N}}, \text{ em que:}$$

$$T = t^3 - t;$$

- número de observações com valores empatados em um grupo de escores empatados.

$\sum T$: somatório de T sobre todos os grupos de escores empatados.

Como exemplo, considere os seguintes postos (apresentados entre parênteses) atribuídos a quatro tratamentos repetidos cinco vezes cada um, em delineamento inteiramente casualizado:

T₁	T₂	T₃	T₄
2,1 (9)	1,9 (7)	2,5 (13)	2,8 (16)
2,0 (8)	2,4 (12)	2,6 (14)	3,0 (18)
1,8 (6)	1,7 (5)	1,4 (2)	2,9 (17)
2,3 (11)	1,6 (4)	2,7 (15)	3,1 (19)
2,2 (10)	1,5 (3)	1,3 (1)	3,2 (20)
R ₁ = (44)	R ₂ = (31)	R ₃ = (45)	R ₄ = (90)

Sendo N = 20, tem-se a seguinte estimativa para H:

$$\hat{H} = \frac{12}{20(21)} \left[\frac{(44)^2}{5} + \frac{(31)^2}{5} + \frac{(45)^2}{5} + \frac{(90)^2}{5} \right] - 3(21) = 0,0286 [2604,4] - 63 = 11,41$$

Considerando um nível de significância $\alpha = 0,05$ e sendo $k - 1 = 3$ graus de liberdade, tem-se que o χ^2 tabelado equivale a 7,82. Como $\hat{H} > \chi^2_3$, rejeita-se a hipótese de nulidade H_0 e conclui-se pela existência de diferenças significativas entre tratamentos.

É importante dizer que quando $k = 2$, deve-se empregar o **teste bilateral de Wilcoxon**, denominado teste da soma das ordens, que é adequado para a comparação de duas amostras independentes.

Em termos de poder-eficiência, a prova de Kruskal-Wallis tem eficiência assintótica de $3/\pi = 95,5\%$ em relação à mais poderosa prova paramétrica, o teste F aplicado em condições em que as suposições básicas da análise de variância são satisfeitas.

Para comparações múltiplas não paramétricas são utilizadas estatísticas que são funções de $|R_i - R_j|$ (contraste entre tratamentos dois a dois) e referem-se a complementações do próprio teste de Kruskal-Wallis. Campos (1983) apresenta com detalhes esses procedimentos.

13.4.2. Análise de variância de Friedman para o delineamento em blocos casualizados

A prova de Friedman (Friedman, 1937) é adequada ao caso de delineamentos em classificação cruzada ou dupla, ou seja, em situações de ensaios em blocos casualizados. O método apresenta as mesmas suposições e exigências da prova de Kruskal-Wallis e também é um substituto do teste F para testar a hipótese da nulidade dos efeitos de k tratamentos.

Nesta prova, os dados são dispostos em uma tabela de dupla entrada com b linhas (blocos) e k colunas (tratamentos) e o método consiste dos seguintes passos:

- a) Ordenamento, por linha, dos tratamentos dos blocos, de 1 a k.
- b) Determinação da soma de postos por coluna, ou seja, para cada um dos k tratamentos, sendo esta soma denominada R_j .
- c) Cômputo da estatística de Friedman, dada por:

$$V = \frac{12}{bk(k+1)} \sum_{i=1}^k R_j^2 - 3b(k+1);$$

- d) verificação da significância de V, a qual, sob H_0 e para amostras grandes tem distribuição χ^2 com $k - 1$ graus de liberdade.

Quando ocorrem empates entre observações de um mesmo bloco, atribui-se a média dos postos a cada um deles e aplica-se a seguinte correção (apenas quando a proporção de empates é alta) em V:

$$V' = \frac{V}{C}, \text{ em que:}$$

$$C = 1 - \frac{\sum_i T_i}{bk(k^2 + 1)}$$

$$T_i = \sum_j t_{ij}^3 - k$$

t_{ij} : número de escores empatados no grupo j do bloco i.

Considere o mesmo exemplo apresentado no tópico anterior, agora tratando-o no delineamento em blocos casualizados. Assim:

Tratamentos				
Blocos	T ₁	T ₂	T ₃	T ₄
B ₁	2,1 (2)	1,9 (1)	2,5 (3)	2,8 (4)
B ₂	2,0 (1)	2,4 (2)	2,6 (3)	3,0 (4)
B ₃	1,8 (3)	1,7 (2)	1,4 (1)	2,9 (4)
B ₄	2,3 (2)	1,6 (1)	2,7 (3)	3,1 (4)
B ₅	2,2 (3)	1,5 (2)	1,3 (1)	3,2 (4)
	R ₁ = (11)	R ₂ = (8)	R ₃ = (11)	R ₄ = (20)

A partir deste quadro, o teste de Friedman é aplicado de forma similar a um teste F usado nos postos de observações nos tratamentos, em vários blocos.

Sendo b = 5 e k = 4, tem-se a seguinte estimativa para V:

$$V = \frac{12}{5(4)(5)} (11^2 + 8^2 + 11^2 + 20^2) - 3(5)(5) = 0,12(706) - 75 = 9,72$$

Considerando uma aproximação para amostras grandes, $\alpha = 0,05$ e sendo $k - 1 = 3$ graus de liberdade, tem-se que o χ^2 tabelado equivale a 7,82. Como $V > \chi^2_3$, rejeita-se a hipótese H_0 , de igualdade das médias de tratamentos.

Quando $k = 2$, deve-se empregar o **teste bilateral do Sinal**, o qual é adequado para a comparação de duas amostras pareadas. O teste de Friedman equivale a uma generalização do teste do Sinal para $k > 2$ tratamentos.

Quanto ao poder do teste de Friedman, comparado com a prova paramétrica mais poderosa (teste F), os resultados apresentados por Friedman (1937) são muito favoráveis a esta prova.

Para comparações múltiplas não paramétricas também são utilizadas estatísticas que são funções de $|R_i - R_j|$ (contraste entre tratamentos dois a dois) as quais referem-se a complementações do teste de Friedman. Para maiores detalhes, sugere-se consultar a obra de Campos (1983).

13.5. Estimadores Melhorados para a Média de Tratamentos

Os experimentos relativos a tratamentos silviculturais são geralmente delineados com os objetivos de comparação desses tratamentos e de inferência sobre as suas médias. Em tais experimentos, os tratamentos são realizados sobre materiais genéticos que podem ser mais ou menos homogêneos dependendo das condições e objetivos da pesquisa. Os efeitos dos tratamentos podem ser influenciados pelos erros aleatórios a que está sujeita a pesquisa e pelos efeitos genotípicos particulares de cada indivíduo que recebe o tratamento. Assim, esse tópico tem por objetivo apresentar estimadores para médias de tratamentos, que forneçam informações livres das influências genotípicas e dos erros aleatórios.

Considerando o delineamento experimental de blocos ao acaso, com t tratamentos, b blocos e n plantas por parcela (plantas obtidas aleatoriamente de uma população), tem-se o modelo matemático:

$$Y_{ijk} = \mu + t_i + b_j + e_{ij} + d_{ijk},$$
 onde μ , t_i , b_j , e_{ij} e d_{ijk} referem-se aos efeitos da média geral, de tratamentos, de blocos, de parcelas e de indivíduos dentro de parcela, respectivamente.

Considerando t_i , e_{ij} e d_{ijk} como efeitos aleatórios e b_j como efeito fixo (essas suposições não alteram os resultados da análise de variância e das comparações das médias pelo teste F, conforme Steel & Torrie, 1980), para efeito dessa apresentação, tem-se a seguinte distribuição dos efeitos genotípicos (G) por meio dos efeitos do modelo estatístico:

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

$$b_1 = \frac{\sigma_G^2}{\sigma_d^2} = \frac{h_G^2}{1-t_1} = \frac{h_G^2}{1-c^2};$$

$$b_2 = \frac{\sigma_G^2/n}{\frac{\sigma_d^2}{n} + \sigma_e^2} = \frac{h_G^2}{1+(n-1)t_1} = \frac{h_G^2}{1+(n-1)c^2};$$

$$b_3 = \frac{\sigma_G^2/nb}{\frac{\sigma_d^2}{nb} + \frac{\sigma_e^2}{b} + \sigma_t^2} = \frac{h_G^2}{1+(n-1)t_1 + (nb-1)t_2} = \frac{h_G^2}{1+(n-1)c^2 + (nb-1)t_i}, \text{ em que:}$$

$t_1 = c^2 = \frac{\sigma_e^2}{\sigma_F^2}$ = correlação intraclasse entre indivíduos na parcela ou correlação devida ao ambiente comum da parcela.

$t_2 = t_t = \frac{\sigma_t^2}{\sigma_F^2}$ = correlação intraclasse entre os indivíduos que recebem um mesmo tratamento.

h_G^2 = herdabilidade individual no sentido amplo

$\sigma_F^2 = \sigma_t^2 + \sigma_e^2 + \sigma_d^2$ = variância fenotípica.

Verifica-se pelas expressões de b_1 , b_2 e b_3 que a maioria da variação genotípica presente no material experimental fica retida nos efeitos de indivíduo dentro de parcela ($Y_{ijk} - \bar{Y}_{ij.}$) e de parcela ($\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}$). Apenas a fração σ_G^2/nb permanece confundida nos efeitos de tratamentos.

Com base neste resultado, pode-se dizer que, com nb muito baixo, os valores genotípicos das plantas que recebem os tratamentos podem mascarar os efeitos de tratamento, exceto quando o material experimental for homogêneo ($\sigma_G^2 = h_G^2 = 0$) ou a herdabilidade no sentido amplo for muito baixa.

O uso de material experimental homogêneo só é desejável quando o objetivo for a realização de recomendações específicas para o material em experimentação. Quando o objetivo for a extração dos resultados para outros materiais genéticos é recomendável a utilização de material genético heterogêneo e, neste caso, com um número razoável de plantas para cada tratamento.

O efeito de tratamento ($\bar{Y}_{i..} - \bar{Y}_{...}$) pode ser assim decomposto:

$$\begin{aligned} (\bar{Y}_{i..} - \bar{Y}_{...}) &= \left(\frac{nbt_t}{1+(n-1) c^2 + (nb-1) t_t} + \frac{h_G^2}{1+(n-1) c^2 + (nb-1) t_t} + \right. \\ &\quad \left. + \frac{nc^2}{1+(n-1) c^2 + (nb-1) t_t} + \frac{(1-h_G^2 - t_t - c^2)}{1+(n-1) c^2 + (nb-1) t_t} \right) (\bar{Y}_{i..} - \bar{Y}_{...}) \\ &= \left(\frac{nb t_t + h_G^2 + nc^2 + (1-h_G^2 - t_t - c^2)}{1+(n-1) c^2 + (nb-1) t_t} \right) (\bar{Y}_{i..} - \bar{Y}_{...}) \end{aligned}$$

Assim, o efeito de tratamento, na realidade, é devido a quatro fatores:

- (i) Efeito do tratamento propriamente dito, que é função de nbt_t .
- (ii) Efeito genotípico, que é função de h_G^2 .
- (iii) Efeito ambiental entre parcelas, que é função de nc^2 .
- (iv) Efeito ambiental dentro de parcelas, que é função de $(1-h_G^2 - t_t - c^2)$.

Para a detecção de diferenças significativas entre tratamentos é desejável que o efeito descrito em (i) seja muito superior aos demais e que explique a quase totalidade de ($\bar{Y}_{i..} - \bar{Y}_{...}$). Uma ação que pode permitir atingir tal objetivo é aumentar o número de repetições, pois, neste caso, aumenta-se nbt_t sem aumentar nc^2 e reduzindo, também, (σ_G^2/nb) .

Com base no exposto, tem-se que a fração de ($\bar{Y}_{i..} - \bar{Y}_{...}$) que realmente determina a média do tratamento é:

$$R^2 = \frac{nb t_t}{1+(n-1) c^2 + (nb-1) t_t} = \frac{\sigma_t^2}{\frac{\sigma_d^2 + \sigma_e^2}{nb} + \sigma_t^2} = 1 - 1/F, \text{ onde } F \text{ é o valor do teste F}$$

de Snedecor da análise de variância.

Assim, a determinação (R^2) associada ao efeito de tratamento pode ser generalizada para efeitos fixos ou aleatórios de tratamentos, pois depende apenas de F , que é idêntico para ambos os modelos. A média (valor esperado em uma nova repetição do ensaio ou em uma aplicação do tratamento em termos operacionais) verdadeira (M) do tratamento é então, melhor predita fazendo-se $M = (1-1/F)(\bar{Y}_{i..} - \bar{Y}_{...}) + \bar{Y}_{...} = (1-1/F)\bar{Y}_{i..} + (1/F)\bar{Y}_{...}$ (Resende, 1999). Em outras palavras, a melhor estimativa da média é dada pela fração (1-1/F) da média amostral do tratamento mais a fração (1/F) da média geral. Esta expressão para a média, além de ser mais coerente com a prática, considera adequadamente a precisão experimental na estimativa da média de tratamento, pois quanto mais preciso (> F) o experimento, menor é a correção (-1/F) aplicada sobre as médias amostrais de tratamento. Neste caso, níveis de significância diferentes (por exemplo 5% e 1%) conduzirão a médias ajustadas diferentes, embora na metodologia convencional conduzissem a inferências idênticas sobre as médias.

Considere o exemplo apresentado por Steel & Torrie (1980), referente à avaliação do teor de óleo de sementes de linho, cujas plantas foram inoculadas em seis diferentes estádios de crescimento. Estes seis tratamentos foram avaliados no delineamento de blocos casualizados com quatro repetições, produzindo os seguintes resultados:

Fontes de Variação	GL	SQ	QM	E (QM)	F
Blocos	3	3,14	1,05	-	
Tratamentos	5	31,65	6,33	$\sigma_d^2 / n + \sigma_e^2 + b\sigma_t^2$	4,83**
Resíduo	15	19,72	1,31	$\sigma_d^2 / n + \sigma_e^2$	

** significativo para $p<0,01$ de probabilidade de erro Tipo I.

	Tratamentos	Médias
	1	5,1
	2	4,3
	3	4,0
	4	6,7
	5	6,0
	6	7,0
	Média Geral	5,517

Supondo diferentes valores de F, ter-se-iam as seguintes inferências sobre as médias:

Trata- mentos	M origi- nal	M com $F \rightarrow \infty$ (P valor $\rightarrow 0$)	M com $F = 4,56$ (P valor $= 0,01$)	M com $F = 2,90$ (P valor $= 0,05$)	M com $F = 2,27$ (P valor $= 0,10$)	M com $F \rightarrow 1,00$
1	5,1	5,1	5,2	5,2	5,3	5,517
2	4,3	4,3	4,6	4,7	4,8	5,517
3	4,0	4,0	4,3	4,5	4,7	5,517
4	6,7	6,7	6,4	6,3	6,2	5,517
5	6,0	6,0	5,9	5,8	5,7	5,517
6	7,0	7,0	6,8	6,5	6,3	5,517

Constata-se que a média amostral equivale ao valor mais provável da média, apenas quando $F \rightarrow \infty$ ou seja, quando P valor $\rightarrow 0$. Com a redução de $F \rightarrow \infty$ para $F = 1$, o valor mais provável da média converge para a média geral e a variância entre as médias converge para zero. No caso, a inferência pontual

mais correta para a média do melhor tratamento (tratamento 6) é 6,7 (visto que o valor de F obtido foi de 4,83 e o p valor 0,01) e não 7,0. Verifica-se, também, que a diferença absoluta entre o melhor e o pior tratamento cai de 3,0 para 2,4, reduzindo-se em 20%. É importante relatar que o cômputo do P-valor só é válido quando os efeitos de tratamentos são fixos. Neste tópico a apresentação do P-valor foi apenas ilustrativa.

Do ponto de vista da estatística clássica ou freqüentista, a média amostral é o melhor estimador não viciado (BLUE) da média populacional. O estimador apresentado, que mostra um regressor (no caso 1-1/F) da média amostral do tratamento em relação à média geral é viciado (Gianola & Fernando, 1986). Entretanto, existem estimadores e preditores viciados que propiciam menor erro quadrático médio do que os estimadores BLUE (Henderson, 1984), sendo assim vantajosos em relação aos estimadores não viciados (Efron, 1975).

Os três principais procedimentos de estimação caracterizam os estimadores: melhor estimador linear não viciado (BLUE); estimador de mínimo erro quadrático médio (EMMSE) e estimador Bayesiano (EB). Esses três estimadores são similares e idênticos sob certas circunstâncias (Henderson, 1984).

A utilização e o reconhecimento das vantagens dos estimadores viciados iniciou-se com o trabalho de Stein (1955) que constituiu um verdadeiro paradoxo na Estatística. Stein demonstrou que a média aritmética é estimador não admissível, isto é, que existem estimadores que propiciam menor erro quadrático médio ou menor risco que a média aritmética. Neste contexto, James & Stein (1961) apresentaram um estimador melhorado para a média populacional, o qual é dado por $M^* = k (\bar{Y}_{..} - \bar{Y}_{...}) + \bar{Y}_{...}$, em que k é um fator regressor da média amostral de determinado tratamento sobre a média geral. Constata-se, assim, a similaridade entre M e M^* .

O fator k apresentado por James & Stein é dado por $K = 1 - \frac{(t-3) \sigma^2}{\sum (\bar{Y}_{..} - \bar{Y}_{...})^2}$, em que t refere-se ao número de tratamentos e σ^2 , à variância residual. Notando a similaridade entre $(t - 3)$ e o número de graus de liberdade $(t - 1)$ para tratamentos, verifica-se também grande similaridade entre k e $(1 - 1/F)$. Stein (1955) demonstrou que a média aritmética é um estimador admissível apenas quando existem uma ou duas médias a serem estimadas, de forma que a correção é necessária somente quando três ou mais tratamentos forem considerados, conforme intuitivamente esperado. O estimador melhorado apresentado por James & Stein (1961) não necessita de qualquer suposição referente aos efeitos fixos ou aleatórios ou distribuições das médias a serem estimadas (Efron & Morris, 1977). Os estimadores de mínimo erro quadrático médio têm recebido mais atenção recentemente (Bibby & Toutenburg, 1977).

O procedimento de estimação bayesiana é mais antigo (1763) do que o método de Stein e também minimiza o erro quadrático esperado, de forma que o estimador de James-Stein é muito similar ao estimador de Bayes, tornando-se idênticos para grande número de tratamentos (Efron & Morris, 1977), sendo

denominados de estimadores de Bayes-Stein, Bayes empírico ou regra empírica de Bayes. Em inferência bayesiana não existe qualquer distinção entre efeitos fixos ou aleatórios, e os parâmetros a serem estimados são considerados variáveis aleatórias (Gianola & Fernando, 1986) que devem ser estimadas considerando as incertezas associadas a elas. No caso de inferências sobre médias populacionais de tratamentos, sob o enfoque bayesiano, Box & Tiao (1973) apresentam como regressor exatamente a quantidade $(1 - 1/F)$.

Com base no exposto, recomenda-se, para inferências (com menor risco) práticas na área de pesquisa agropecuária e florestal, a apresentação das médias de tratamento como $M = (1 - 1/F) \bar{Y}_{i..} + (1/F)\bar{Y}_{...}$, em que F , $\bar{Y}_{i..}$ e $\bar{Y}_{...}$ referem-se ao valor de F para tratamentos na análise de variância, média de tratamentos e média geral, respectivamente. A validade deste procedimento tem sido comprovada na prática (obtendo-se maior coincidência entre os valores reais dos tratamentos em termos operacionais e aqueles estimados com base neste procedimento) e também por meio de simulações (obtendo-se maior coincidência entre os valores paramétricos e aqueles estimados com base neste método).

13.6. Teste de Efeitos Fixos e Aleatórios no Contexto dos Modelos Mistas

13.6.1. Teste de efeitos fixos

A análise de dados de experimentos balanceados é tradicionalmente realizada por uma análise de variância na qual a soma de quadrados total dos desvios das observações em relação às suas médias é partitionada em soma de quadrados devidas aos efeitos de tratamentos (efeitos fixos) e a outros efeitos aleatórios, tais como o erro experimental. Sob a hipótese de nulidade (sob H_0) dos efeitos de tratamentos, a soma de quadrados de tratamentos segue uma distribuição múltipla (produto dos graus de liberdade por uma combinação linear dos componentes de variância dos efeitos aleatórios e somatório dos quadrados dos efeitos fixos) de uma χ^2 . A soma de quadrados associada aos efeitos aleatórios (erro, por exemplo), sob as suposições distribucionais padrões, também segue uma distribuição múltipla de uma χ^2 (Elston, 1998).

Uma vez que a razão entre distribuições χ^2 segue uma distribuição F de Snedecor, hipóteses sobre os efeitos de tratamentos podem ser testadas pela razão entre os quadrados médios dos efeitos de tratamentos e do erro (por exemplo), por meio do teste F. Segundo Elston (1998), alguns dos benefícios de um delineamento experimental, ou seja, do balanceamento, referem-se ao fato da variância de todos os contrastes para um dado fator de tratamento ter a mesma esperança sob a hipótese da nulidade e tal variância ser tão pequena quanto possível. Por outro lado, a análise de dados que não apresentam essas propriedades de balanceamento é muito menos fidedigna, e os principais pro-

blemas são os fatos da decomposição da soma de quadrados não ser única e da não existência de um quadrado médio do erro com igual esperança sob a hipótese da nulidade.

Segundo Henderson (1984), os testes de hipóteses referentes aos efeitos fixos são exatos apenas quando os componentes de variância (variância dos efeitos aleatórios) são conhecidos. Isto porque quando se usam as estimativas destes componentes de variância em lugar de seus valores paramétricos, as somas dos quadrados mencionadas anteriormente não têm distribuição χ^2 ou qualquer outra distribuição tratável. Apenas no caso de um modelo de efeitos fixos (exceto o erro), em que a variância do erro –Var (e) – é definida como $\text{Var}(e) = l\sigma_e^2$, as distribuições χ^2 e F mencionadas podem ser asseguradas, quando substitui-se σ_e^2 por sua estimativa $\hat{\sigma}_e^2$ (obtida por máxima verossimilhança restrita – REML). Também, para classificações duplas, com um efeito fixo e outro aleatório, testes exatos para os efeitos fixos podem ser encontrados (Khuri & Littell, 1987; Gallo & Khuri, 1990).

Segundo Henderson (1984), duas possibilidades existem para contornar o problema: (i) estimar os componentes de variância por máxima verossimilhança restrita e invocar o teste da razão de verossimilhança – LRT –, através do qual, sob suposição de normalidade e com amostras grandes, -2 log da razão de verossimilhança aproxima uma distribuição χ^2 ; (ii) considerar os componentes de variância estimados, como sendo paramétricos (isto é tanto mais realístico quanto maior for o tamanho da amostra) e realizar um teste aproximado usando as distribuições χ^2 (para os quadrados médios) e F (para a razão entre os quadrados médios e consequentemente para teste dos efeitos fixos). Nos dois casos, surge a questão do que significa uma amostra grande com dados desbalanceados. Ainda, segundo Henderson (1984), certamente N (o número total de dados) tendendo ao infinito não é uma condição suficiente para se ter uma grande amostra. Considerações a respeito do número de níveis em cada um dos efeitos aleatórios e da proporção de subclasses perdidas são importantes para responder essa questão. Consequentemente, a validade de uma aproximação qui-quadrado para o LRT e do próprio uso de uma χ^2 (para os quadrados médios) é incerta.

Henderson Júnior & Henderson (1979) consideraram quatro alternativas computacionais para a realização de testes de hipóteses com dados desbalanceados: (i) tratar os efeitos aleatórios como fixos; (ii) ignorar os efeitos aleatórios; (iii) tratar alguns efeitos aleatórios como fixos e ignorar outros; (iv) usar estimativas das variâncias dos efeitos aleatórios nas equações de modelo misto. Para a maioria das situações, tais autores preferem a alternativa (iv). O principal argumento baseia-se na lógica de que, no contexto dos modelos mistos, quando os efeitos aleatórios são tratados como fixos, tais modelos são tratados como se as suas variâncias (dos efeitos aleatórios) fossem de magnitudes muito altas em relação à variância do erro. Quando os efeitos aleatórios são ignorados, eles são tratados como se tivessem variância nula. A realidade, provavelmente, compreende uma situação intermediária entre estes dois extremos.

Do que foi exposto, torna-se claro o problema do teste dos efeitos fixos quando os componentes de variância são estimados a partir de dados desbalanceados. Mais recentemente, Berk (1987) propôs uma estatística para o teste simultâneo de vários fatores de efeitos fixos. Esta é denominada estatística de Wald (Wald, 1943), a qual é uma generalização da estatística T^2 de Hotelling (similar multivariado do teste t de Student) sem a necessidade de a matriz de variância-covariância seguir uma distribuição Wishart. A estatística W de Wald é dada por $W = (\hat{\beta} - \beta_o)' \hat{\Sigma} (\hat{\beta} - \beta_o)$ em que β_o é o vetor paramétrico de efeitos fixos sob $H_0: \beta = \beta_o$, $\hat{\beta}$ é a estimativa de β_o e $\hat{\Sigma}$ é a estimativa da matriz de informação de Fisher avaliada em $\hat{\beta}$.

Assintoticamente, a estatística de Wald tem distribuição χ^2 , ou seja, $(\hat{\beta} - \beta_o)' \hat{\Sigma} (\hat{\beta} - \beta_o) \sim \chi^2$. Assim, as estatísticas LRT e W são assintoticamente equivalentes e, sob H_0 , convergem em distribuição para uma variável χ^2 , sendo, entretanto, o LRT o critério que define um teste uniformemente mais poderoso (Silvey, 1975).

No entanto, os diferentes contrastes de efeitos fixos têm variâncias desiguais, cujas aproximações à distribuição χ^2 têm diferentes números de graus de liberdade. Assim, os graus de liberdade do denominador (erro) não são facilmente determinados na análise de um modelo misto desbalanceado. Segundo Gilmour et al. (2000), somente em poucos casos bem definidos, podem ser, precisamente, determinados os graus de liberdade do erro. A estatística de Wald distribuída como uma χ^2 considera que a variância do erro é conhecida, ou seja, assume infinitos graus de liberdade, sendo por isto mais frouxa.

Em resumo, na prática três estatísticas têm sido recomendadas para teste dos efeitos fixos: (i) LRT (Welham & Thompson, 1997); (ii) W de Wald (Kenward & Roger, 1997; Elston, 1998); (iii) F de Snedecor (Gilmour et al., 2000). Segundo pesquisas recentes, a estatística de Wald para pequenas amostras é aproximada por uma distribuição F como uma razão entre as variáveis aleatórias χ^2 . Assim, embora outras estatísticas possam ser introduzidas para o teste de efeitos fixos, a estatística de Wald é atrativa porque ela reproduz exatamente a análise de variância para delineamentos balanceados.

Em função do exposto, no software ASREML (Gilmour et al., 2000), os testes de hipóteses referentes aos efeitos fixos no contexto da estimação por REML têm sido baseados em tabelas de análise de variância (contemplando apenas os efeitos fixos) usando a própria estatística F de Snedecor. Em grandes análises, por exemplo, em modelos com grande número de efeitos aleatórios, os números de graus de liberdade serão altos e os testes estatísticos não serão muito sensíveis (Kennedy, 1991) a pequenas diferenças nos graus de liberdade do denominador. Segundo Gilmour et al. (2000) é preferível o teste F, em que se pode ter ao menos uma indicação (baseada na estrutura dos dados) de qual o número de graus de liberdade deve ser usado.

No contexto do melhoramento genético, pode haver o interesse em testar os efeitos fixos de blocos, por exemplo, visto que a diferença significativa entre blocos é desejável. Sendo significativo o efeito de blocos, os materiais genéticos são melhor testados para condições diversas de ambiente, permitindo a seleção de materiais mais estáveis, ou seja, ocorre uma redução nos efeitos da interação genótipo x ambiente em relação à seleção baseada no comportamento em vários blocos similares.

13.6.2. Teste de efeitos aleatórios

O uso da tabela de análise de variância para a construção de testes F para os efeitos aleatórios em modelos desbalanceados é muito difícil. Isto porque é necessária a obtenção dos quadrados médios a partir dos componentes de variância e seus multiplicadores, multiplicadores estes, muito difíceis de serem computados sob desbalanceamento. O software Genstat (Cullis et al., 1999) apresenta, de forma aproximada, esta opção.

Existe uma maneira mais formal para testar os efeitos aleatórios, ou seja, para verificar se determinado efeito aleatório necessita permanecer no modelo. Essa abordagem formal baseia-se no teste LRT. Um procedimento prático é descrito por Gilmour et al. (2000). Tendo-se a estimativa do componente de variância ($\hat{\sigma}^2$) do referido efeito aleatório e seu respectivo desvio padrão ($\hat{s}(\hat{\sigma}^2)$) pode-se concluir:

- (i) se $\hat{\sigma}^2 - 2(\hat{s}(\hat{\sigma}^2)) > 0$ (ou alternativamente $\hat{\sigma}^2 / (\hat{s}(\hat{\sigma}^2)) > 1$), o efeito aleatório é significativo e deve permanecer no modelo. Essa assertiva é válida e conservadora pois:
 - se $\hat{\sigma}^2 / (\hat{s}(\hat{\sigma}^2)) > 2$: o LRT será significativo.
 - se $\hat{\sigma}^2 / (\hat{s}(\hat{\sigma}^2)) < 0,5$: o LRT será não significativo.
 - se $1 < \hat{\sigma}^2 / (\hat{s}(\hat{\sigma}^2)) < 2$: existe a necessidade de aplicar o LRT.

A expressão $\hat{\sigma}^2 - 2(\hat{s}(\hat{\sigma}^2)) > 0$ indica, com confiança maior do que 95%, que o referido componente de variância, não apresenta valor paramétrico igual a zero.

- (ii) se $\hat{\sigma}^2 - 2\hat{s}(\hat{\sigma}^2) < 0$ (ou $\hat{\sigma}^2 / (\hat{s}(\hat{\sigma}^2)) < 1$), o efeito aleatório é não significativo e pode ser retirado do modelo.

Para ser um pouco mais exato, deve-se calcular $\hat{\sigma}^2 / (\hat{s}(\hat{\sigma}^2))$ e aplicar o LRT quando $1 < \hat{\sigma}^2 / (\hat{s}(\hat{\sigma}^2)) < 2$. A aplicação do LRT é apresentada com mais detalhes no Capítulo 9. Tal teste envolve duas vezes a redução no Log L resultante da retirada de t termos aleatórios, quantidade esta distribuída como uma χ_t^2 . Assim, para a verificação da significância de um efeito aleatório tem-se que

$LRT \sim \chi^2_1$. Entretanto, Stram & Lee (1994) sugerem uma correção por meio da multiplicação do P valor associado a χ^2_1 por 0,5, ou seja, sugerem o uso de uma distribuição $\chi^2_{0,5}$. Esta correção é, sobretudo, indicada para teste no limite do espaço paramétrico, quando o P valor aproximado para a estatística de teste d (duas vezes a redução no Log L) é $0,5(1-P(\chi^2_1 \leq d))$, em que P denota probabilidade.

É importante ressaltar que os Log L residuais não são comparáveis quando mudam-se os efeitos fixos ou a parte fixa do modelo.

13.7. Análise de Variância Multivariada

Muitas vezes, na pesquisa agropecuária, o interesse na comparação de tratamentos leva em consideração uma série de variáveis simultaneamente. Este é um caso em que vetores médios relativos a vários caracteres são comparados e é uma extensão ou generalização das comparações de médias, empregadas nas análises univariadas.

Dentre as situações mais importantes encontram-se a comparação entre dois vetores médios, por meio do teste T^2 de Hotelling, e a comparação de vários vetores médios, pela análise de variância multivariada (Manova) empregando-se o teste Λ de Wilks. Esses procedimentos são apresentados com detalhes por Johnson & Wickern (1988).

13.7.1. Teste T^2 de Hotelling

Para a comparação de duas médias, no caso univariado, tem sido comum o uso do teste t, cuja estatística de teste é dada por $t = \frac{\bar{y} - \mu_o}{s / \sqrt{n}} \sim t_{n-1}$, em que \bar{y} é a média amostral, s é o desvio padrão amostral e n o número de observações por tratamento. Neste caso, testa-se a hipótese $H_o: \mu = \mu_o$. Expressando esse teste de forma quadrática, $t^2 = \frac{(\bar{y} - \mu_o)^2}{s^2 / n} = n (\bar{y} - \mu_o)(s^2)^{-1}(\bar{y} - \mu_o) \sim t_{n-1}^2 = F_{1, n-1}$.

No caso multivariado, tem-se a extensão do teste t, de forma que a generalização do quadrado da distância $(\bar{y} - \mu_o)$ fornece $T^2 = n (\bar{y} - \mu_o)'(\bar{S})^{-1}(\bar{y} - \mu_o)$, em que:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \bar{y}_i ; S = \frac{1}{n-1} \sum_{i=1}^n (\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})'; \mu_o = [\mu_{o1} \mu_{o2} \cdots \mu_{op}]$$

T^2 é denominada estatística T^2 de Hotelling, a qual tem distribuição

$$T^2 \sim \frac{(n-1)p}{n-p} F_{p, n-p}$$

Para comparação de dois vetores de médias de dois tratamentos, o teste da hipótese $H_0: \mu_1 = \mu_2$ é feito aplicando-se a estatística T^2 de Hotelling na forma:

$$T^2 = [(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)]' \left[S_p \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{-1} [(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)].$$

Neste caso, T^2 tem distribuição F de Snedecor, ou seja,

$$T^2 \frac{(n_1 + n_2 - p - 1)}{(n_1 + n_2 - 2) p} \sim F_{p, n_1 + n_2 - p - 1}.$$

Assim, rejeita-se H_0 se $T^2 \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2) p} > F_{p, n_1 + n_2 - p - 1}(\alpha)$, em que p refere-se ao número de variáveis.

Como exemplo, considere a avaliação de dois tratamentos (1 e 2) com os seguintes vetores de médias e matrizes de desvios padrões para duas variáveis:

$$\bar{y}_1 = \begin{bmatrix} 8,3 \\ 4,1 \end{bmatrix}; \quad \bar{y}_2 = \begin{bmatrix} 10,2 \\ 3,9 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}; \quad S_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

Sendo $n_1 = n_2 = 50$, então:

$$S_p = \frac{(n_1 - 1) S_1 + (n_2 - 1) S_2}{n_1 + n_2 - 2} = \frac{49 \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix} + 49 \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}}{98} = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}$$

Sendo $H_0: \mu_1 - \mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, a estimativa de T^2 é:

$$T^2 = \left(\begin{bmatrix} 8,3 & 10,2 \\ 4,1 & 3,9 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)' \left(\begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix} \left(\frac{1}{50} + \frac{1}{50} \right) \right)^{-1} \left(\begin{bmatrix} 8,3 & 10,2 \\ 4,1 & 3,9 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = 52,47$$

Em consequência:

$$T^2 \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2) p} = \frac{52,47 (50 + 50 - 2 - 1)}{(50 + 50 - 2) 2} = 2597$$

Sendo $F_{2,97}^{(0,05)} = 3,1$, rejeita-se H_0 (pois, $25,97 > 3,1$) e conclui-se pela existência de diferença bivariada significativa entre os tratamentos.

13.7.2. Teste Λ de Wilks (Manova)

Para comparação entre vetores médios de vários tratamentos, uma técnica corrente é à análise de variância multivariada – Manova. Associado à Manova, aplica-se o teste Λ de Wilks (lambda de Wilks), o qual é análogo ao teste F da análise de variância univariada.

Na aplicação da Manova, as suposições quanto à estrutura e distribuição dos dados são:

- (i) independência: as amostras aleatórias dos diferentes tratamentos são independentes.
- (ii) homocedasticidade: todos os tratamentos tem matriz de covariância Σ .
- (iii) normalidade multivariada: cada tratamento é normal p-variado.

Esta última condição pode ser relaxada quando os tamanhos amostrais são grandes (ou seja, quando o teorema central do limite se aplica). De maneira geral, a técnica da análise de variância multivariada é robusta a pequenos desvios das suposições. A seguir tal procedimento é descrito.

Modelo

$$\underset{\sim}{y}_{ij} = \underset{\sim}{\mu} + \underset{\sim}{\tau}_i + \underset{\sim}{\varepsilon}_{ij}, \quad i=1, 2, \dots, g \text{ e } j=1, 2, \dots n_i, \text{ em que:}$$

$\underset{\sim}{\mu}$: vetor médio global.

\sim

$\underset{\sim}{\tau}_i$: efeito do iésimo tratamento.

\sim

$\underset{\sim}{\varepsilon}_{ij}$: vetor de erros, aleatórios e independentes ($[\underset{\sim}{\varepsilon}_{ij} \sim N(\underset{\sim}{0}, \underset{\sim}{\Sigma})]$).

Soma dos Quadrados (SQ)

$$\begin{aligned} SQ \ Total &= \sum_{i=1}^g n_i (\underset{\sim}{\bar{y}}_i - \underset{\sim}{\bar{y}}_.) (\underset{\sim}{\bar{y}}_i - \underset{\sim}{\bar{y}}_.)' + \sum_{i=1}^g \sum_{j=1}^{n_i} (\underset{\sim}{\bar{y}}_{ij} - \underset{\sim}{\bar{y}}_{i.}) (\underset{\sim}{\bar{y}}_{ij} - \underset{\sim}{\bar{y}}_{i.})' \\ &= SQ \ Tratamento + SQ \ Erro \end{aligned}$$

$$= B + W$$

A soma de quadrados do erro pode ser escrita como:

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{\bar{y}}_{..}) (\bar{y}_{ij} - \bar{\bar{y}}_{..})' = (n_1 - 1) S_1 + (n_2 - 1) S_2 + \dots + (n_{p-1}) S_p.$$

Hipóteses

$$H_0: \tau_1 = \tau_2 = \dots = \tau_g = 0$$

H_1 : pelo menos um dos τ_i é diferente dos outros.

Manova

Fontes de Variação	GL	Matriz das somas de quadrados e produtos cruzados	Λ de Wilks
Tratamentos	$g-1$	$B = \sum_{i=1}^g n_i (\bar{y}_i - \bar{\bar{y}}_{..}) (\bar{y}_i - \bar{\bar{y}}_{..})'$	$\Lambda = \frac{ W }{ B+W }$
Erro	$\sum_{i=1}^g n_i - g$	$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{\bar{y}}_{..}) (\bar{y}_{ij} - \bar{\bar{y}}_{..})'$	$\Lambda = \frac{ W }{ B+W }$
Total	$\sum_{i=1}^g n_i - 1$	$B+W$	

O teste Λ de Wilks envolve variâncias generalizadas, em que se rejeita H_0 quando a razão de variância generalizada $\Lambda = \frac{|W|}{|B+W|}$ é muito pequena.

Para amostras grandes, $-(n-1-\frac{p+g}{2}) \ln \Lambda \sim \chi^2_{p(g-1)}$ e, para $g \geq 2$ tratamentos e $p = 2$ variáveis, $\sum_{i=1}^g \frac{n_i - g - 1}{g - 1} \frac{1 - (\Lambda)^{1/2}}{(\Lambda)^{1/2}} \sim F_{p(g-1), p(n-g-1)}$.

Como exemplo, considere a avaliação de três tratamentos (1, 2 e 3) para duas variáveis, com os seguintes vetores de médias:

$$T_1: \begin{bmatrix} 9 \\ 3 \end{bmatrix}; \quad \begin{bmatrix} 6 \\ 2 \end{bmatrix}; \quad \begin{bmatrix} 9 \\ 7 \end{bmatrix}; \quad \Rightarrow \bar{\bar{y}}_{..} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

$$T_2: \begin{bmatrix} 0 \\ 4 \end{bmatrix}; \quad \begin{bmatrix} 2 \\ 0 \end{bmatrix}; \quad \Rightarrow \bar{\bar{y}}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$T_3: \begin{bmatrix} 3 \\ 8 \end{bmatrix}; \quad \begin{bmatrix} 1 \\ 9 \end{bmatrix}; \quad \begin{bmatrix} 2 \\ 7 \end{bmatrix}; \quad \Rightarrow \bar{\bar{y}}_3 = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$$

Sendo $n_1=3$; $n_2=2$ e $n_3 = 3$, tem-se:

$$\bar{y}_{..} = \begin{bmatrix} \frac{3x8+2x1+3x2}{3+2+3} \\ \frac{3x4+2x2+3x8}{3+2+3} \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

Assim:

$$B = \sum_{i=1}^g n_i (\bar{y}_i - \bar{y}_{..}) (\bar{y}_i - \bar{y}_{..})' = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$$

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{y}_i) (\bar{y}_{ij} - \bar{y}_i)' = \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$$

A Manova resultante apresenta os seguintes resultados:

Fontes de Variação	GL	Soma de quadrados e produtos
Tratamentos	$g-1 = 3-1 = 2$	$B = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$
Erro	$n-g = 8-3 = 5$	$W = \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$
Total	$n-1 = 8-1 = 7$	$B+W = \begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$

A estimativa do lambda de Wilks equivale a:

$$\Lambda = \frac{|W|}{|B+W|} = \frac{\begin{vmatrix} 10 & 1 \\ 1 & 24 \end{vmatrix}}{\begin{vmatrix} 88 & -11 \\ -11 & 72 \end{vmatrix}} = \frac{239}{6215} = 0,0385$$

Para verificar a significância de Λ , obtém-se:

$$\frac{\sum_{i=1}^g n_i - g - 1}{g - 1} \frac{1 - (\Lambda)^{1/2}}{(\Lambda)^{1/2}} = \frac{8 - 3 - 1}{3 - 1} \frac{1 - (0,0385)^{1/2}}{(0,0385)^{1/2}} = 8,20$$

Sendo $F_{p(g-1), p(n-g-1)(\alpha)} = F_{4,8}^{(0,05)} = 3,84$ e a magnitude da estatística $(8,20)$ maior que o valor crítico $(3,84)$, rejeita-se a hipótese H_0 de que os tratamentos são iguais.

Para verificar quais tratamentos diferem significativamente, podem ser utilizados os intervalos de Bonferroni (Johnson & Wichern, 1988).

No melhoramento de plantas perenes, a análise de variância multivariada apresenta grande relevância no contexto dos experimentos com medidas repetidas. Os tradicionais modelos de repetibilidade e de análise de variância univariada baseiam-se na suposição de correlação genética igual a 1 (requisito para o uso do modelo de repetibilidade) e correlação fenotípica constante para todos os pares de idades ou medidas (requisitos para o uso dos modelos de repetibilidade e de análise de variância univariada).

Em realidade, as correlações entre medidas repetidas realizadas em intervalos de tempo mais próximos tendem a ser maiores do que entre medidas realizadas a um maior intervalo de tempo. Este fato aponta que, nestas situações, o uso de um modelo multivariado tende a ser mais eficiente. Dias & Resende (2001a) discutem a aplicação de um teste estatístico (teste de esfericidade) na avaliação da estrutura da matriz de covariância (ou correlação) entre medidas repetidas, para inferir sobre as eficiências relativas da análise de variância univariada, análise de variância multivariada e análise de variância univariada ajustada (ajustando o P-valor associado à rejeição de H_0 tendo por base o teste F univariado).

Ainda com relação a medidas repetidas, há outras abordagens, tendo em vista que o fator tempo é quantitativo, podendo gerar respostas polinomiais. Dentro destas, citam-se as técnicas da regressão aleatória, do ajuste de *splines* e também da análise espacial via modelos auto-regressivos e/ou de médias móveis. Em realidade, vários modelos geoestatísticos e de análise de séries temporais podem ser aplicados na análise de dados temporalmente dependentes.

13.8. Análise de Tabelas de Contingência

A comparação entre tratamentos para variáveis categóricas ou qualitativas envolve, via de regra, a análise de tabelas de contingência. Um dos principais testes aplicáveis à análise destas tabelas é o qui-quadrado (χ^2). O teste χ^2 tem sido aplicado, desde o início do século passado, ao estudo dos caracteres genéticos qualitativos, por meio da comparação entre freqüências observadas de determinada classe de uma variável e freqüências esperadas, com base em uma proporção teórica, no caso, as freqüências mendelianas. Este é o teste “ χ^2 de ajuste” ou “ χ^2 de aderência” (conforme apresentado no tópico referente ao teste de normalidade dos dados), o qual verifica o ajuste ou a aderência de um conjunto de dados a uma distribuição teórica.

No entanto, o teste χ^2 pode também ser utilizado para estudar a dependência (correlação, conforme usado no Capítulo 8 para a estimativa de repetibilidade) entre duas variáveis categóricas ou entre efeitos de linhas e colunas (“ χ^2 de independência”), visando à comparação entre dois tratamentos ou amo-

tras quanto a uma variável categórica (" χ^2 de homogeneidade ou heterogeneidade") e, também, no estudo de caracteres quantitativos categóricos, ou seja, na análise de vários tratamentos para variáveis categóricas, por meio do teste " χ^2 de heterogeneidade".

O teste χ^2 de homogeneidade (também denominado de heterogeneidade) para comparação entre dois tratamentos baseia-se em tabelas de contingência 2 x 2, conforme apresentado a seguir.

Tabela 13.2. Tabela de contingência 2 x 2.

Tratamentos	Categorias		Totalis
	Presença	Ausência	
Tratamento 1	n_{11}	n_{12}	$n_{1.}$
Tratamento 2	n_{21}	n_{22}	$n_{2.}$
	$n_{..1}$	$n_{..2}$	$n_{..}$

A estatística qui-quadrado, neste caso, é dada por $\hat{\chi}^2 = \sum_{ij=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, em que

O_{ij} e E_{ij} referem-se às freqüências observadas e esperadas em cada uma das c caselas da tabela. A freqüência esperada na casela ij é dada por $E_{ij} = \frac{(n_{i.})(n_{.j})}{n_{..}}$ e a

freqüência observada nesta mesma casela é o próprio n_{ij} . Assim, o valor observado na casela 11 é n_{11} e o valor esperado para esta casela é $E_{11} = \frac{(n_{1.})(n_{.1})}{n_{..}} = \frac{(n_{11} + n_{12})(n_{11} + n_{21})}{n_{..}}$.

Substituindo-se estes desdobramentos na expressão de χ^2 , obtém-se, para o caso de tabelas de contingência 2 x 2:

$$\hat{\chi}^2 = \frac{n_{..} (n_{11} n_{22} - n_{21} n_{12})^2}{(n_{1.})(n_{.2})(n_{1.})(n_{2.})}$$

Por exemplo, sendo $n_{11} = 23$, $n_{12} = 52$, $n_{21} = 14$ e $n_{22} = 6$, tem-se:

$\hat{\chi}^2 = \frac{95 (23*6 - 14*52)^2}{(37)(58)(75)(20)} = 10,27$. Este mesmo resultado também pode ser obtido computando-se separadamente as freqüências esperadas. Neste caso, $E_{11} = 29,21$; $E_{12} = 45,79$; $E_{21} = 7,79$; $E_{22} = 12,21$ e $\hat{\chi}^2 = \sum_{ij=1}^4 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 1,32 + 0,84 + 4,95 + 3,16 = 10,27$.

A estatística $\hat{\chi}^2$, quando H_0 é verdadeira, aproxima uma distribuição χ^2 com 1 grau de liberdade. Assim, para um nível de significância α , rejeita-se a hipótese H_0 de homogeneidade quando $\hat{\chi}^2 \geq \chi_{\alpha,1}^2$, em que $\chi_{\alpha,1}^2$ é o valor tabelado da distribuição qui-quadrado com 1 grau de liberdade e nível de significância α . Para melhorar a aproximação (sob H_0) de $\hat{\chi}^2$ a uma distribuição qui-quadrado, recomenda-se a correção de continuidade de Yates, especialmente em caso de valores esperados muito pequenos (neste caso, recomenda-se também o uso de um outro teste, o teste exato de Fisher). Com a correção de continuidade de Yates, a estatística qui-quadrado é dada por $\hat{\chi}^2 = n.. (\ln_{11} n_{22} - n_{21} n_{12}) / 0,5 n.. / ((n_{1.})(n_{2.})(n_{1.})(n_{2.}))$.

A análise de tabelas de contingência com vários (t) tratamentos é uma extensão da análise apresentada para tabelas 2×2 . Nesta situação, têm-se as tabelas $t \times 2$, quando a variável apresenta apenas 2 categorias ou $t \times v$, quando a característica é avaliada com base em várias categorias. Neste caso, usa-se a estatística do χ^2 de heterogeneidade, conforme apresentado a seguir.

Considere-se a avaliação de 10 tratamentos (progênies) (no delineamento inteiramente casualizado com 18 plantas por tratamento), para o caráter sobrevivência, em que as plantas vivas receberam o escore 1 e as mortas receberam o escore zero. Os dados encontram-se resumidos na Tabela 13.3.

Tabela 13.3. Tabela de contingência 10×2 , referente à avaliação de 10 tratamentos (progênies) para o caráter sobrevivência, segundo duas categorias (plantas vivas ou mortas).

Tratamento	Freqüências	Vivas	Mortas	Total	$\hat{\chi}_1^2$	P-valor
1	O	18	0	18	2,1242	0,10 < p < 0,20
	E	16,1	1,9	18		
2	O	14	4	18	2,5950	0,10 < p < 0,20
	E	16,1	1,9	18		
3	O	12	6	18	2,8914	0,001 < p < 0,01
	E	16,1	1,9	18		
4	O	18	0	18	2,1242	0,10 < p < 0,20
	E	16,1	1,9	18		
5	O	16	2	18	0,0059	0,90 < p < 0,95
	E	16,1	1,9	18		
6	O	17	1	18	0,4766	0,30 < p < 0,50
	E	16,1	1,9	18		
7	O	17	1	18	0,4766	0,30 < p < 0,50
	E	16,1	1,9	18		
8	O	18	0	18	2,1242	0,10 < p < 0,20
	E	16,1	1,9	18		
9	O	15	3	18	0,7120	0,30 < p < 0,50
	E	16,1	1,9	18		
10	O	16	2	18	0,0059	0,90 < p < 0,95
	E	16,1	1,9	18		
Total		O	161	19	180	0,0000
		E	161	19	180	

Os valores esperados foram computados empregando-se o estimador E_{ij} apresentado anteriormente e o $\hat{\chi}^2$ em cada tratamento, bem como o total, foram estimados por $\hat{\chi}^2 = \sum_{ij=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$.

O $\hat{\chi}^2_{t-1}$ de heterogeneidade é dado por:

$$\hat{\chi}^2_{t-1} = \sum \hat{\chi}^2_i - \hat{\chi}^2_T = (2,1242 + 2,5950 + \dots + 0,0059) - 0,0000 = 20,5360 - 0,0000 = 20,5360,$$

em que $\hat{\chi}^2_T$ é o qui-quadrado do total (Tabela 13.3).

O valor estimado $\hat{\chi}^2_9 = 20,560$ comparado com um valor tabelado (por exemplo 19,68 a 2% de significância) de uma distribuição qui-quadrado com 9 graus de liberdade, indica a presença de heterogeneidade, ou seja, de diferenças significativas entre tratamentos, para a variável sobrevivência. O próprio $\hat{\chi}^2_1$ para a progénie 3 revela também esta heterogeneidade, visto que a sua frequência observada diferiu significativamente da esperada, fato que não ocorreu para as demais progénies.

No caso de tabelas $t \times 2$, ou seja, quando a variável avaliada pode assumir apenas 2 categorias, o cálculo de $\hat{\chi}^2_{t-1}$ é bastante simplificado, podendo ser feito pela expressão:

$$\hat{\chi}^2_{t-1} = \left(\sum_{i=1}^t \frac{y_i^2}{N_i} - \frac{(\sum y_i)^2}{\sum N_i} \right) / (\hat{p}\hat{q}), \text{ em que:}$$

y_i : número de indivíduos que receberam o escore 1 (plantas vivas, no caso) no tratamento i .

N_i : número de indivíduos avaliados no tratamento i .

$\hat{p} = \frac{\sum_{i=1}^t y_i}{\sum_{i=1}^t N_i}$: estimativa da incidência média da característica no experimento (fração de plantas vivas no caso).

$$\hat{q} = 1 - p.$$

No presente caso, $y_i = 18, 14, \dots, 15, 16$; $\sum y_i^2 = 2627$,

$$\sum y_i = 161; N_i = 18; \sum N_i = 180; \hat{p} = \frac{161}{180} = 0,8944 \text{ e } \hat{q} = 0,1056 \text{ Assim,}$$

sendo N_i constante para todos os tratamentos:

$$\hat{\chi}^2_9 = \frac{1}{N_i} \sum_{i=1}^t y_i^2 - \frac{(\sum y_i)^2}{\sum N_i} = \frac{1}{18} * 2627 - \frac{(161)^2}{180} = 20,5360$$

O χ^2 de heterogeneidade pode também ser utilizado para o estudo de caracteres quantitativos, inclusive para a estimativa da herdabilidade. Considerando o efeito de progénies de meios irmãos como aleatório, uma estimativa da herdabilidade (\hat{h}_a^2) na escala binomial pode ser obtida no presente exemplo:

$$\hat{h}_a^2 = [4[\hat{\chi}_{t-1}^2 - (t-1)]]/[N(t-1)] = 4[20,5360 - 9]/[18(9)] = 0,285.$$

Esta estimativa da herdabilidade individual no sentido restrito é equivalente àquela obtida via análise de variância dos dados, no delineamento inteiramente casualizado, conforme apresentado a seguir:

FV	GL	SQ	QM	E(QM)
Tratamentos	9	1,9389	$Q_1 = 0,2154$	$\sigma^2 + N\sigma_t^2$
Erro	170	15,0556	$Q_2 = 0,0886$	σ^2
Total	179	16,9944		

Têm-se as estimativas:

$$\hat{\sigma}^2 = Q_2 = 0,0886 \quad \sigma_t^2 = (Q_1 - Q_2)/N = (0,2154 - 0,0886)/18 = 0,0070$$

$$\hat{h}_a^2 = \frac{4 * \hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}^2} = \frac{0,028}{0,0956} = 0,29.$$

Esta coincidência de resultados é esperada quando não existem efeitos fixos no modelo, além da média geral. É importante verificar que o numerador de $\hat{\chi}_9^2$ equivale à própria soma de quadrados de tratamentos e o denominador equivale, aproximadamente, (difere por arredondamento) à variância fenotípica ($\hat{\sigma}_t^2 + \hat{\sigma}^2$).

Capítulo 14

**Delineamento de Experimentos
para a Estimação de Parâmetros
Genéticos e Seleção**

14.1. Introdução

Estimativas confiáveis de parâmetros genéticos são necessárias nos procedimentos de predição de valores genéticos. A rigor, para a avaliação genética, são necessárias estimativas fidedignas de parâmetros genéticos das populações-base (originais) não selecionadas e não endogâmicas.

Definida a população de referência e o parâmetro de interesse, a escolha de um delineamento ótimo para a estimação precisa dos componentes de variância genética e fenotípica deve basear-se no critério de minimização da variância da estimativa do referido parâmetro. Tal critério conduz à determinação do tamanho amostral adequado para a estimação dos parâmetros com determinada precisão, escolhida a priori. Abordagens relevantes associadas a este tema foram apresentadas por Robertson (1959), Thompson (1989) e Hill (1990). Aspectos práticos da estimação de parâmetros genéticos em espécies perenes foram relatados também por Resende & Rosa-Perez (1999a).

No melhoramento de espécies perenes, os estudos sobre o número adequado de repetições e plantas por parcela na experimentação têm enfatizado a detecção de diferenças significativas entre tratamentos (Lee, 1983; Lambeth et al., 1983; Cotterill & James, 1984 e Haapanen, 1992), a eficiência na seleção de progêneres (Castro, 1992) e de indivíduos, genitores e clones (Resende, 1995).

14.2. Minimização da Variância da Estimativa da Herdabilidade Baseada na Covariância de Meios-irmãos

A variância da estimativa da herdabilidade individual no sentido restrito, obtida com base na covariância de meios-irmãos, equivale a (Robertson, 1959)

$$\text{Var}(h_a^2) = 32[1 + (N-1) h_a^2 / 4] (1 - h_a^2 / 4)^2 / [N(N-1)(p-1)] \text{ ou, aproximadamente,}$$

$$\text{Var}(h_a^2) = 32 h_a^2 / Np, \text{ em que:}$$

N : número de indivíduos por famílias de meios-irmãos

p : número de famílias de meios-irmãos

Para um número total (Np) fixo de indivíduos (tomados aleatoriamente na população não selecionada e com iguais números de indivíduos por família), a variância é minimizada (por diferenciação de $\text{Var}(h_a^2)$) quando $N = 4/h_a^2 + 1$ ou, aproximadamente, $N = 4/h_a^2$ (Robertson, 1959). Assim, o número adequado de indivíduos por família é da ordem de 80, 40, 27, 20 e 13 para herdabilidades da ordem de 0,05; 0,10; 0,15; 0,20 e 0,30, respectivamente. O número N adequado é um compromisso entre estimar precisamente a variância dentro de famílias (N grande) e a variância entre famílias (N pequeno). Os números N recomendados são menores do que aqueles necessários para a maximização da acurácia seletiva (Resende, 1995). Uma vez que os dados experimentais são, em geral, utilizados tanto para o propósito de estimação de parâmetros genéticos, quanto para a seleção, os valores de N utilizados na prática são maiores que os reco-

mendados. Este fato não é problemático devido ao fato de discrepâncias de N em relação ao ótimo terem pequeno efeito na eficiência da estimação, visto que o delineamento de meios-irmãos é robusto, exceto quando as discrepâncias forem para valores menores do que o indicado (Thompson, 1989).

O número total $T = Np$ de indivíduos pode ser determinado escolhendo-se a priori um valor desejado do desvio padrão da herdabilidade, sendo dado por $T = [32 h_a^2 / \text{Var}(h_a^2)]^{1/2}$.

Na Tabela 14.1 são apresentadas as magnitudes adequadas de T , N e p para alguns valores de herdabilidade e desvios padrões desejados.

Observa-se que, para obter o desvio padrão equivalendo a 50% do valor da herdabilidade, em caracteres de baixa herdabilidade ($< 0,30$), o número adequado de indivíduos é relativamente pequeno e aumenta à medida que a h_a^2 diminui. Para obter um valor fixo de $\text{Var}(h_a^2)^{1/2}$, caracteres com maiores herdabilidades exigem maiores números de indivíduos e de famílias em relação aos caracteres com mais baixa h_a^2 . Para obter $\text{Var}(h_a^2)^{1/2}$ igual a 20% das estimativas, N da ordem de 2.600 a 8.000 são necessários para h_a^2 de 0,30 a 0,10, respectivamente.

Tabela 14.1. Números adequados de famílias de meios-irmãos (p), de filhos por genitor ou tamanho de famílias de meios-irmãos (N) e população experimental total ($T = Np$) para obter estimativas de herdabilidade (h_a^2) com um determinado desvio padrão ($[\text{Var}(h_a^2)]^{1/2}$).

h_a^2	Situação	$[\text{Var}(h_a^2)]^{1/2}$	T	N	p
0,05		0,025	2560	80	32
0,10	Desvio	0,050	1280	40	32
0,15	Padrão	0,075	853	27	32
0,20	Equivalendo	0,100	640	20	32
0,25	a 50% de h_a^2	0,125	512	16	32
0,30		0,150	426	13	32
0,05		0,05	640	80	8
0,10		0,05	1280	40	32
0,15	Valor fixo de	0,05	1920	27	72
0,20	$\text{Var}([h_a^2])^{1/2}$	0,05	2560	20	128
0,25	igual a 0,05	0,05	3200	16	200
0,30		0,05	3840	13	295
0,05		0,01	16000	80	200
0,10	$\text{Var}(h_a^2)^{1/2}$	0,02	8000	40	200
0,15	Igual	0,03	5333	27	200
0,20	a 20% de h_a^2	0,04	4000	20	200
0,25		0,05	3200	16	200
0,30		0,06	2666	13	200

Quanto à precisão (desvio padrão) necessária nos procedimentos de previsão de valores genéticos, tem sido verificado que o uso de índices combinados baseados em informação de família são robustos a erros na herdabilidade, visto que pouco progresso genético é perdido quando se usam herdabilidades incorretas. Já índices baseados em vários caracteres são menos robustos (Sales & Hill, 1976 a e b; Thompson, 1989). Uma alternativa poderia ser a consideração dos intervalos de confiança na faixa de 10% (desvios padrões de 0,05) para as h_a^2 em que os intervalos de 0,05 a 0,15; 0,10 a 0,20; 0,15 a 0,25; 0,20 a 0,30 e 0,25 a 0,35 conteriam as herdabilidades de 0,10; 0,15; 0,20; 0,25 e 0,30, respectivamente. Neste caso, os números de indivíduos adequados estariam na faixa de 1.200 a 3.800, conforme o terço médio da Tabela 14.1. Para obter desvios padrões equivalentes a 20% das estimativas, N da ordem de 2.600 a 8.000 são necessários, para herdabilidades de 0,30 a 0,10 , respectivamente, conforme o terço inferior da Tabela 14.1.

Considerando N fixo igual a 1.000, seriam obtidos os seguintes intervalos de confiança: 0,01 a 0,09; 0,04 a 0,16; 0,08 a 0,22 e 0,12 a 0,28 para herdabilidades de 0,05; 0,10; 0,15 e 0,20 e p de 13; 25; 37 e 50 famílias, respectivamente. Assim, pode-se inferir que pelo menos cerca de 1.000 dados e mais do que 20 famílias deveriam ser utilizados na estimação de parâmetros genéticos. O número de 1.000 foi recomendado também como valor mínimo para estimação de parâmetros genéticos com vistas à construção de índices de seleção relativos a caracteres com herdabilidade menores ou igual a 0,20 (Harris, 1964).

Resultados experimentais sobre a herdabilidade em acácia-negra são apresentados a seguir. O estudo foi realizado a partir de experimentação realizada com 43 progênieis de meios-irmãos oriundas de populações naturais de *Acacia mearnsii* existentes na Austrália.

O delineamento estatístico utilizado na instalação do experimento foi o bloco ao acaso, com 10 repetições, sendo cada parcela representada por seis árvores, plantadas linearmente. Aos 3 anos de idade foi mensurado o diâmetro das árvores, à altura do peito (DAP).

As estimativas das variâncias e herdabilidades foram obtidas por Resende et al. (1998a) pelo método iterativo da máxima verossimilhança restrita –REML – utilizando o programa DFREML (Meyer, 1998). Os componentes de variância e parâmetros genéticos foram estimados a partir de subconjuntos de dados de diferentes tamanhos utilizando-se 3 (6 combinações diferentes), 6 (4 combinações diferentes) e 10 blocos.

Em estudos realizados no Brasil e na China, com progênieis de *A.mearnsii*, as estimativas da herdabilidade individual para o caráter DAP variaram de 0,29 a 0,37, com média 0,33 (Tabela 14.2).

Tabela 14.2. Estimativas de herdabilidade para o caráter diâmetro (DAP) obtidas de literatura.

Autor	Número de Progênes	Número de Repetições	Árvores por parcela	Idade	DAP (h^2 individual)
Resende et al. 1992	46	5	5	3	0,31
Resende et al. 1992	46	5	5	3	0,36
Bi et al., 1991	169	20	1	2	0,29
Bi et al., 1991	169	20	1	2	0,37
Média					0,33

Aos 3 anos de idade, a sobrevivência das árvores nos dez blocos foi de 81,3%, ou seja, das 2.580 mudas plantadas, 2.099 foram avaliadas. Na Tabela 14.3 são apresentadas as estimativas dos parâmetros genéticos e fenotípicos associadas ao experimento de acácia-negra.

Tabela 14.3. Valores estimados de média, variâncias, herdabilidade e desvio padrão da herdabilidade em função do número de repetições utilizados.

Blocos	T	Média	Desvio Padrão	Variância Fenotípica	Variância Aditiva	h_a^2	[Var(h_a^2)] $^{1/2}$
1 a 10	2099	8.02	1.99	3.8627	1.2981	0.3345	0,0714
1, 2, 3, 4, 5 e 6	1269	7.99	2.00	3.9530	1.4431	0.3651	0,0959
1, 2, 4, 5, 7 e 8	1284	8.05	1.96	3.8268	1.5564	0.4067	0,1006
4, 5, 6, 7, 8 e 9	1257	8.10	1.98	3.7795	1.0328	0.2733	0,0834
2, 3, 5, 6, 8 e 9	1262	8.13	2.00	3.8479	0.9264	0.2408	0,0781
Média	1268	8.07	1.99	3.8518	1.2397	0.3215	0.0895
CV%	1%	1%	1%	2%	25%	24%	12%
1, 2 e 3	633	7.98	1.99	3.9760	1.3500	0.3394	0,1309
4, 5 e 6	636	8.00	2.01	3.9305	0.9337	0.2376	0,1093
7, 8 e 9	621	8.20	1.95	3.6284	0.8293	0.2285	0,1085
1, 4 e 7	628	7.94	1.94	3.8242	1.7999	0.4707	0,1548
2, 5 e 8	656	8.15	1.97	3.8216	1.1925	0.3121	0,1233
3, 6 e 9	606	8.09	2.04	3.9026	1.0573	0.2709	0,1196
Média	629	8.06	1.98	3.8472	1.1938	0.3099	0.1244
CV%	3%	1%	2%	3%	29%	29%	14%

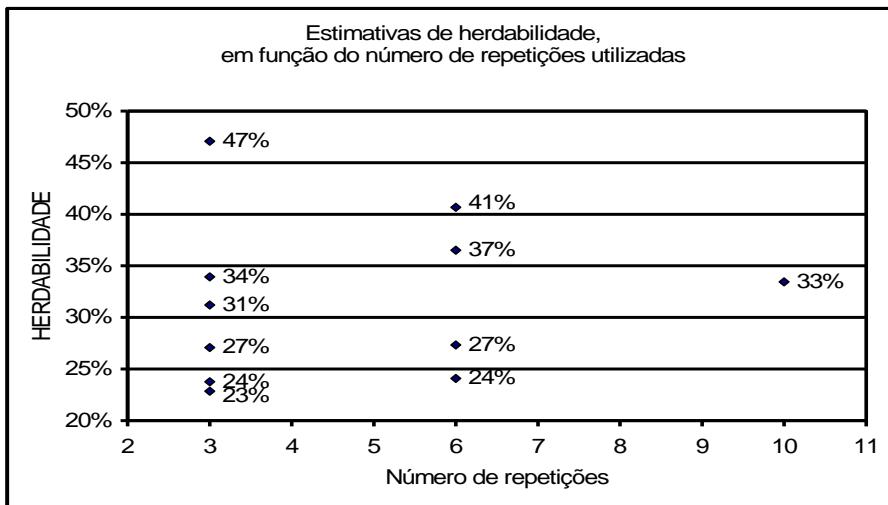


Figura 14.1. Valores de herdabilidades em função do número de repetições analisadas.

Verifica-se que a média e a variância fenotípica foram bem estimadas com 3, 6 ou 10 repetições. As estimativas da variância aditiva e da herdabilidade oscilaram bastante quando foram utilizadas 3 e 6 repetições (Tabela 14.3). Constatase, também, que os desvios-padrões da h^2 foram muito maiores (proporcionalmente) quando obtidos com apenas 3 repetições, conforme esperado teoricamente (Tabela 14.1).

A estimativa mais fidedigna da herdabilidade foi de $0,33 \pm 0,07$, obtida com base em 2.099 plantas (10 repetições) das 43 progénies. O valor médio das estimativas obtidas com base em 6 repetições (em torno de 1.268 plantas) foi de 0,32 e a média das estimativas obtidas com base em 3 repetições (em torno de 629 plantas) foi de 0,31. Considerando a média das estimativas obtidas de literatura (0,33) e os resultados alcançados no presente estudo, pode-se inferir que um valor plausível da herdabilidade para o DAP em acácia-negra situa-se ao redor de 0,30.

Com base nos resultados da Tabela 14.3 e no gráfico da Figura 14.1, pode-se inferir que estimativas individuais da h_a^2 obtidas com base em 3 e 6 repetições são inadequadas, pois conduziriam a valores bastante discrepantes do valor mais provável. Dessa forma, para a estimativa fidedigna da herdabilidade devem ser considerados (pelo menos aproximadamente) os tamanhos amostrais recomendados. Considerando o tamanho de progénie geralmente adotado na prática do melhoramento florestal (6 plantas por parcela e 10 repetições), números esses que maximizam a acurácia seletiva, conforme Resende (1995), são apresentados na Tabela 14.4 os tamanhos amostrais recomendados para a estimativa da herdabilidade, bem como o número de famílias associado.

Os valores de T mostrados na Tabela 14.4 reforçam a inferência de que quando o número total de dados for inferior a 1.000, não se recomenda utilizá-

los para a estimação de parâmetros genéticos. Neste caso, a predição de valores genéticos deve ser realizada utilizando-se valores médios das herdabilidades obtidas na literatura. Mc Cutchan et al.(1989) também concluíram que ao menos 1.000 observações, associadas a uma sobrevida de pelo menos 80%, são necessárias para a estimativa da h_a^2 .

Tabela 14.4. Tamanho amostral (T) aproximado, número de plantas por família (N = 60) e número de famílias (p) adequados à estimativa da herdabilidade (h_a^2) com razoável precisão ($s(h_a^2)$).

h_a^2	$(s(h_a^2))$	T	N	p
0,05	0,01	16.000	60	267
0,10	0,02	8.000	60	133
0,15	0,03	5.400	60	90
0,20	0,04	4.000	60	67
0,25	0,05	3.200	60	53
0,30	0,06	2.600	60	43
0,40	0,08	2.000	60	33
0,50	0,10	1.600	60	27
0,60	0,12	1.400	60	23

Para a estimativa da herdabilidade no sentido amplo (h_g^2), tem-se que a variância de h_g^2 , estimada pela análise de correlação intraclasse, é dada por $Var(h_g^2)=8 h_g^2 / N$. Verifica-se por esta expressão que, para herdabilidades e precisões similares, a estimativa de h_g^2 requer menor número (N) total de indivíduos. Com N = 15 repetições ou ramos por clone, pode-se inferir que são necessários 17 clones para que h_g^2 seja estimada com razoável precisão (similar àquela obtida com N = 1.000 indivíduos em estrutura de famílias de meios-irmãos, para a estimativa de h_a^2 no sentido restrito).

Com base no exposto, pode-se concluir que:

- (i) Para uma precisa estimativa da herdabilidade no sentido restrito, a partir da avaliação de progêniens de meios-irmãos, recomendam-se os tamanhos amostrais aproximados de 8.000, 4.000 e 2.600 dados para herdabilidades de 0,10; 0,20 e 0,30, respectivamente.
- (ii) Para estimativa da média e variância fenotípica de um caráter quantitativo, a utilização de um tamanho amostral equivalente a cerca de 1/3 do tamanho amostral adequado à estimativa da variância aditiva e da herdabilidade já é suficiente para a obtenção de estimativas precisas.

- (iii) Com tamanho amostral inferior a 1.000, não se recomenda a estimativa de parâmetro genéticos. Neste caso, a predição de valores genéticos deve ser realizada, preferencialmente, por meio de estimativas de parâmetros relatadas em literatura.
- (iv) Para a estimativa adequada da herdabilidade no sentido amplo, pelo menos 250 ramegos (por exemplo 17 clones com 15 ramegos de cada) são necessários.

14.3. Estimação da Herdabilidade Via Regressão Pais-filhos

Este delineamento baseia-se em parentes não colaterais, pertencentes a duas gerações. Os filhos podem ser progêneros de meios-irmãos ou irmãos germanos. Sendo σ_x^2 e σ_y^2 as variâncias fenotípicas nos genitores e progêneros, respectivamente, t a correlação fenotípica intraclasse dos membros da família na progênie, r a correlação entre genitores (em número igual a p) e progêneros e β o coeficiente de regressão, tem-se, para famílias de tamanho N :

$$V(\beta) = [(t - r^2) + (1-t)/N] (\sigma_y^2 / p) \sigma_x^2$$

Diferenciando a expressão de $V(\beta)$, Latter & Robertson (1960), obtiveram o tamanho ótimo de família $N = [(1-t)/(t-r^2)]^{1/2}$. Assim, para $t = 0,25$ h_a^2 e $\beta = h_a^2/2 = r$, são encontrados os valores de N iguais a 9, 7, 5 e 4 para herdabilidades de 0,05; 0,10; 0,20 e 0,40, respectivamente. Verifica-se, então, que os tamanhos da família adequados à estimativa da herdabilidade por regressão são bem menores que aqueles necessários para estimar o mesmo parâmetro via covariância de irmãos. Assim, com N baixo, é melhor utilizar os métodos de regressão e não os de correlação intraclasse para estimativa da herdabilidade.

Considerando $N = 1$ indivíduo por genitor e a expressão $V(h_a^2) = 4V(\beta)$, são apresentados na Tabela 14.5 estimadores aproximados (Falconer, 1989) para a variância da estimativa de herdabilidade estimada pelo método de regressão, bem como a precisão obtida com determinado número de dados. São apresentados, também, resultados de dois métodos de correlação intraclasse.

Tabela 14.5. Estimadores e estimativas para o desvio padrão da herdabilidade $[V(h_a^2)]^{1/2}$ associados a diferentes métodos de estimativa.

Método	Estimador de $[V(h_a^2)]^{1/2}$	h_a^2	$T = Np$	Estimativa de $[V(h_a^2)]^{1/2}$
Regressão sobre 1 genitor ($N = 1$)	$2/(s)^{1/2} = 2/(T/2)^{1/2}$	Qualquer	500	0,126
Regressão sobre a média dos genitores ($N = 1$)	$(2/s)^{1/2} = (4/T)^{1/2}$	Qualquer	500	0,089
Correlação de meios-irmãos	$(32 h_a^2 / T)^{1/2}$	0,10	500	0,080
Correlação de irmãos germanos*	$(16 h_a^2 / T)^{1/2}$	0,10	500	0,057

* Considerando ausência de dominância e de variância de ambiente comum. Para a regressão, $T = Np + p = 2p$, ou seja $p = T/2$, pois $N = 1$ descendente para cada genitor, $p =$ número de genitores.

Com base nas expressões para o desvio padrão (Tabela 14.5), observa-se que, para o mesmo número total de indivíduos (T), a estimativa via correlação de meios-irmãos é mais adequada que a regressão filhos sobre um genitor, quando $h_a^2 < 0,25$, visto que a eficiência (em termos de razão de variâncias) é dada por $Ef = [32 h^2/T]/8T = 4h_a^2$. Para $h_a^2 > 0,25$, $Ef > 1,0$ e, portanto, a variância das estimativas obtidas via correlação de meios-irmãos é maior, recomendando-se, então, o uso da regressão.

Para estimação de correlações genéticas, os tamanhos ótimos de família são praticamente os mesmos mencionados para estimativa da herdabilidade. Assim, para regressão envolvendo genitores e suas progêniens de meios-irmãos, tamanhos de família da ordem de 6 são adequados para uma ampla gama de caracteres (com variadas herdabilidades).

14.4. Estimação de Correlações Genéticas

Para estimação de correlações genéticas, se ambos os caracteres são avaliados no mesmo indivíduo, suas herdabilidades e as correlações genéticas e ambiental são aproximadamente iguais, o delineamento ótimo equivale àquele para estimação das herdabilidades. Se as herdabilidades diferem, mas as correlações genéticas e fenotípicas são iguais, os tamanhos ótimos de família estão entre: $4/h_{ai}^2$ e $4/h_{a2}^2$, em que h_{ai}^2 é a herdabilidade do caráter $i = 1, 2$. Se as correlações genéticas e ambientais também diferem, maiores tamanhos de família são necessários (Hill, 1990). Robertson (1959) sugere que o tamanho ótimo de família é dado por $N=4(1+r_a^2)^{1/2}/[h_a^2(1-r_a^2)]$, onde r_a é a correlação genética aditiva. Esta expressão considera herdabilidades iguais para os dois caracteres. A expressão do desvio padrão da estimativa da correção genética,

$$\text{dada por } s(r_{xy}) = \left[\frac{(1-r_{xy}^2)^2}{2} \frac{s(h_{ax}^2)}{h_{ax}^2} \frac{s(h_{ay}^2)}{h_{ay}^2} \right]^{1/2} \text{ em que } s(h_{ax}^2) \text{ é o desvio padrão da}$$

estimativa da herdabilidade do caráter x, pode ser usada para inferências mais generalizadas.

A utilização das estimativas de parâmetros genéticos em índices de seleção deve ser realizada de forma criteriosa, sendo recomendável a inspeção da matriz $P^{-1}G$ (em que P refere-se à matriz de covariância fenotípica e G , à matriz de covariância genética entre os caracteres) em termos de seus autovalores. O ideal é que esta matriz seja positiva-definida (isto é, deve apresentar todos os autovalores positivos) e que os autovalores estejam entre 0 e 1 (Cameron, 1997). A probabilidade de que $P^{-1}G$ apresente um autovalor negativo aumenta à medida que reduz o tamanho amostral e aumenta o número de caracteres incluídos no índice.

A estrutura de correlação intraclasse pode também ser usada no contexto dos estudos de interação genótipo x ambiente, em que as famílias são divididas para avaliação em vários locais. Considerando herdabilidades iguais nos dois locais, o tamanho ótimo de família em cada ambiente é $N = 4(1 + r_a^2)^{1/2} / [h_a^2 (1 - r_a^2)]$, em que r_a é a correlação genética de um ambiente para outro. Assim, para valores de r_a de 0,5; 0,7; 0,8 e 0,90, tem-se N igual a $6/h_a^2$; $9,6/h_a^2$; $14,4/h_a^2$ e $28,4/h_a^2$, respectivamente. Assim, verifica-se que quando a correlação genética entre ambientes é alta, o tamanho ótimo de família para estimativa desta correlação deve ser muito maior do que para estimar herdabilidade.

14.5. Delineamento Baseado na Estimação Combinada via Covariância de Irmãos e Regressão Pais-filhos

As estimativas obtidas com base na covariância de irmãos e de pais-filhos são correlacionadas e as duas podem ser ponderadas, obtendo-se uma estimativa única e mais precisa. Para esta ponderação, em geral, utiliza-se o método da máxima verossimilhança – ML– (Hill & Nicholas, 1974).

A combinação das duas fontes de informação, pelo procedimento ML melhora a eficiência da estimativa para herdabilidades intermediárias ($h_a^2 = 0,20$), mas pouco contribui para valores extremos de h_a^2 . Os tamanhos ótimos de família para a estimativa ML são similares àqueles adequados para o método isolado mais eficiente. Assim, para h_a^2 baixas, tamanhos similares àqueles adotados para a estimativa via correlação de meios-irmãos são necessários. Hill (1990) apresenta os seguintes valores ótimos de N para a estimativa ML: 70, 30, 10 e 4, para herdabilidades da ordem de 0,05; 0,10; 0,20 e 0,40, respectivamente. De maneira geral, conclui-se, então, que para um caráter com h_a^2 da ordem de 20%, o melhor procedimento de estimativa é o método combinado baseado em famílias de tamanho 10.

Os ganhos em precisão, pela estimativa combinada, são da ordem de 20% para h_a^2 intermediárias ($h_a^2 = 0,20$). Para valores altos de h_a^2 , a maioria da informação provém da regressão pais-filhos e para valores baixos de h_a^2 , a maioria da informação advém da covariância de irmãos, e, em ambos os casos, o ganho pelo uso da estimativa combinada é da ordem de 5% (Thompson, 1989).

14.6. Estimação da Repetibilidade

Para a repetibilidade (ρ), o desvio padrão pode ser calculado considerando o estimador do desvio padrão do coeficiente de correlação linear simples, já que a

repetibilidade é uma correlação intraclasse. Assim: $\sigma_{\rho} = \frac{(1-\rho)[1+(m-1)\rho]}{[(1/2)m(m-1)(N-1)]^{1/2}}$ para o caso de mais de duas medições (m) e: $\sigma_{\rho} = \frac{1-\rho^2}{(N-1)^{1/2}}$ para o caso de apenas duas medições.

Considerando duas medições, o número (N) de indivíduos necessários para obter uma estimativa de ρ com um desejado desvio padrão σ_{ρ} é dado por: $N = (1-\rho^2)^2 / \sigma_{\rho}^2 + 1$. Na Tabela 14.6 são apresentados os valores adequados de N para alguns valores de ρ e σ_{ρ} .

Tabela 14.6. Valores adequados do número de indivíduos N a serem medidos para a obtenção de estimativas da repetibilidade (ρ) com determinado valor desejado de σ_{ρ} .

ρ	$\sigma_{\rho}=0,2\rho$	N	σ_{ρ}	N
0,10	0,02	2451	0,05	393
0,20	0,04	577	0,05	370
0,30	0,06	231	0,05	332
0,40	0,08	111	0,05	283
0,50	0,10	57	0,05	226
0,60	0,12	29	0,05	165
0,70	0,14	14	0,05	105
0,80	0,16	6	0,05	53
0,90	0,18	3	0,05	16

Considerando um desvio padrão de 0,05, verifica-se que de 100 a 400 indivíduos são suficientes para obter estimativas precisas de repetibilidade que apresentam valores paramétricos entre 0,70 e 0,10, respectivamente.

14.7. Delineamento para Maximização da Acurácia Seletiva

Os números necessários de indivíduos por progénie de meios irmãos para a obtenção de alguns valores de acurácia na predição de valores genéticos de genitores são apresentados na Tabela 14.7, obtidos a partir do estimador da acurácia.

Tabela 14.7. Número (N) adequado de indivíduos por progênies, em função da herdabilidade (h_a^2) do caráter, para a obtenção de alguns valores de acurácia (r_{aa}) na seleção de genitores.

h_a^2	$r_{aa} = 70\%$	$r_{aa} = 80\%$	$r_{aa} = 90\%$	$r_{aa} = 95\%$	r_{aa} para $n = 100$
0,05	76	141	337	732	0,747
0,10	38	70	167	361	0,848
0,15	25	46	110	238	0,892
0,20	19	34	81	176	0,917
0,25	15	27	64	139	0,933
0,30	12	22	53	115	0,944
0,35	10	19	45	97	0,952
0,40	9	16	39	84	0,958
0,45	8	14	34	74	0,963
0,50	7	13	30	65	0,967
0,60	6	10	25	53	0,973
0,70	5	9	21	44	0,977
0,80	4	8	18	38	0,981
0,90	4	7	15	32	0,983

Com herdabilidade acima de 15%, 100 indivíduos por progênies conduzem a acurárias superiores a 90%. Com herdabilidades de 20%, é necessário cerca de 80 indivíduos por progênies para conseguir 90% de precisão na seleção. Já com herdabilidades menores que 35% são necessários mais de 100 indivíduos por progênies para conseguir 95% de acurácia (Tabela 14.7).

Para a seleção de indivíduos, baseada no índice multiefeitos ou BLUP, de uma só geração, os números necessários de indivíduos por família de meios-irmãos, para obtenção de 90% e 95% da acurácia máxima possível, são apresentados na Tabela 14.8.

Tabela 14.8. Número (N) adequado de indivíduos por família, para maximização da acurácia (r_{aa}) na seleção de indivíduos pela seleção combinada ou BLUP envolvendo uma só geração.

h_a^2	$r_{aa} \text{ max}$	90% da $r_{aa} \text{ max}$	95% da $r_{aa} \text{ max}$
0,05	0,53	280	653
0,10	0,55	94	201
0,15	0,58	52	122
0,20	0,61	32	85
0,25	0,63	16	43
0,30	0,66	10	33
0,35	0,68	6	17
0,40	0,71	2	14
0,45	0,73	1	7
0,50	0,76	1	5
0,60	0,80	1	1
0,70	0,85	1	1
0,80	0,90	1	1
0,90	0,95	1	1

Verifica-se que, para seleção de indivíduos, é impossível obter acurácia de 100%. Para valores de herdabilidade (h_a^2) variando de 5% a 90%, a acurácia máxima (com $n \rightarrow \infty$) possível varia de 53% a 95%. Em geral, com h_a^2 acima de 15%, não se justifica o emprego de mais de 50 indivíduos por progénie (Tabela 14.8).

Para a seleção de indivíduos que vão permanecer no plantio, com base em várias medições em cada indivíduo, os números necessários de medições para cada valor de repetibilidade, visando à obtenção de acurácia de 90%, são apresentados na Tabela 14.9.

Tabela 14.9. Número de medições (m) necessárias para a seleção de indivíduos com 90% de acurácia, em função de diferentes valores de repetibilidade (ρ).

P	M
0,10	36
0,20	16
0,30	9
0,40	6
0,50	4
0,60	3
0,70	2
0,80	1
0,90	1

14.8. Repetibilidade e Número de Medições

A eficiência da utilização de mais de uma medição na seleção é dada pela razão entre as acuráncias pela seleção baseada em várias medições e em uma só medição. Para a seleção individual ou massal, a eficiência é dada por $\{m/[1+(m-1)\rho]\}^{1/2}$. A eficiência em termos de ganho genético por unidade de tempo (por ano) é dada pela razão entre as acuráncias seletivas com m e uma medição multiplicada pela razão entre o intervalo de gerações (L) com uma medição e com m medições, ou seja, $(r_{dam}/r_{dai})(L_1/L_m)$. A eficiência anual é dada por $[m/[1+(m-1)\rho]]^{1/2} [L_1/(L_1 + m - 1)]$, para $m \geq 2$. Alternativamente, $EA = EC [L_1/(L_1 + m - 1)]$, para $m \geq 2$, em que EC é a eficiência por ciclo seletivo.

Na Tabela 14.10 são apresentados os valores de eficiência seletiva por ciclo de seleção e anual, para a seleção massal, considerando $L = (6 + m)$ e vários valores de acurácia e número de medições.

Tabela 14.10. Valores de eficiência seletiva por ciclo de seleção (EC) e eficiência anual (EA) para a seleção massal baseada em médias de diferentes números (m) de medições, considerando diferentes valores de repetibilidade (ρ).

ρ	m	EC	EA*
0,90	1	1,00	1,00
	2	1,03	0,90
	3	1,04	0,80
0,85	2	1,04	0,91
	3	1,05	0,82
	4	1,06	0,73
0,80	2	1,05	0,92
	3	1,07	0,83
	4	1,08	0,74
0,75	2	1,07	0,93
	3	1,10	0,85
	4	1,11	0,77
0,70	2	1,08	0,95
	3	1,12	0,87
	4	1,14	0,79
0,65	2	1,10	0,96
	3	1,14	0,89
	4	1,16	0,81
0,60	2	1,12	0,98
	3	1,17	0,91
	4	1,20	0,84
0,55	2	1,14	0,99
	3	1,20	0,93
	4	1,23	0,86
0,50	2	1,15	1,00
	3	1,22	0,95
	4	1,26	0,89
0,45	2	1,17	1,03
	3	1,26	0,98
	4	1,30	0,91
0,40	2	1,20	1,05
	3	1,29	1,00
	4	1,35	0,94
0,35	2	1,22	1,06
	3	1,33	1,05
	4	1,40	0,98
0,30	2	1,24	1,08
	3	1,37	1,06
	4	1,45	1,02

* Para o cômputo de EA, considerou-se $L = (6+m)$.

Na interpretação da Tabela 14.10 deve-se considerar os diferentes objetivos da seleção: (i) para o melhoramento de curto prazo (maximização do ganho na geração atual), é preciso analisar a EC; (ii) para o melhoramento de longo prazo por seleção recorrente deve-se verificar a EA. Considerando uma repetibilidade estimada de 0,75, verifica-se que para o melhoramento no curto prazo não compensa realizar mais do que três medições e, para o melhoramento no longo prazo por seleção massal, o ideal é selecionar com base em apenas uma medição.

Observa-se, também, pela Tabela 14.10 que, com repetibilidade igual a 0,50, a utilização de uma ou duas medições contribui igualmente para o melhoramento no longo prazo. A utilização de duas e três medições só passa a ser vantajosa (EA maior que 1) para o melhoramento no longo prazo quando a repetibilidade for igual ou inferior a 0,45 e 0,35, respectivamente. As eficiências seletivas por ciclo de seleção (EC) para diferentes métodos de seleção são apresentadas na Tabela 14.11.

Tabela 14.11. Eficiência da utilização de 2 (EC_2), 3 (EC_3) e 4 (EC_4) medições em cada indivíduo em relação à utilização de apenas uma, visando ao melhoramento de um caráter quantitativo com repetibilidade de 0,65 e com herdabilidades no sentido restrito (h_a^2) variáveis.

Método de Seleção	h_a^2	EC_2	EC_3	EC_4
Massal (Individual)	-	1,10	1,14	1,16
Seleção de genitores	0,10	1,04	1,05	1,06
pelo comportamento	0,20	1,02	1,03	1,04
das progênie*	0,30	1,02	1,02	1,03
	0,40	1,01	1,02	1,02
Seleção combinada	0,10	1,05	1,08	1,09
com base em família	0,20	1,05	1,08	1,09
de meios-irmãos*	0,30	1,06	1,08	1,10
	0,40	1,07	1,09	1,11

* Considerando $n = 60$ indivíduos por progénie.

Constata-se que o ganho em eficiência com a utilização de várias medições depende do método de seleção e da herdabilidade do caráter, mesmo com um valor fixo e conhecido da repetibilidade. A eficiência é máxima para os métodos que utilizam apenas a informação do indivíduo e decresce, rapidamente, para os métodos que enfatizam mais a informação de médias de família. Assim, o uso de duas medições propicia 10% a mais (em relação a uma medição) de ganho genético para as modalidades de seleção massal e praticamente não conduz a maior eficiência para a seleção de genitores. Para os métodos que

empregam a informação do indivíduo e da família, o ganho em eficiência reduz-se à metade (5%) ou menos (Tabela 14.11).

Em relação à herdabilidade, a eficiência diminui com o aumento da herdabilidade para os métodos que utilizam apenas as informações de médias e aumenta com o aumento da herdabilidade para os métodos que empregam tanto a informação de médias quanto a informação de indivíduos (Tabela 14.11). Neste último caso, quanto maior a h_a^2 maior é o peso dado à informação de indivíduo e, portanto, maior a eficiência.

Os resultados revelam a necessidade de considerar tanto o método de seleção quanto as herdabilidades dos caracteres na recomendação do número adequado de medições em cada indivíduo, e não apenas a repetibilidade.

14.9. Transformações de Herdabilidades

Em muitos trabalhos científicos são apresentadas estimativas de herdabilidades apenas ao nível de médias (h_{am}^2) de famílias de meios-irmãos ou de irmãos germanos. Tendo em vista que a herdabilidade individual (h_{ai}^2) é mais útil, esta pode ser obtida a partir de h_{am}^2 por:

(a)
$$h_{ai}^2 = \frac{h_{am}^2}{1+(N-1) 0,25 (1-h_{am}^2)}$$
, quando h_{am}^2 provém de famílias de meios-irmãos (seleção entre famílias sem sementes remanescentes).

(b)
$$h_{ai}^{2*} = \frac{h_{am}^2}{1+(N-1) 0,5 (1-h_{am}^2)}$$
, quando h_{am}^2 provém de famílias de irmãos germanos.

Esta última expressão é válida considerando dominância nula ou baixa.

Com famílias de irmãos germanos, em presença de dominância, a h_i^{2*} assim estimada contém o verdadeiro h_{ai}^2 mais $(h_{gi}^2 - h_{ai}^2)/2$, ou seja, $h_{ai}^{2*} = h_{ai}^2 + (h_{gi}^2 - h_{ai}^2)/2 = (h_{gi}^2 + h_{ai}^2)/2$, onde h_{gi}^2 é a herdabilidade individual no sentido amplo.

(c)
$$h_{ai}^2 = \frac{4 h_{am}^2}{h_{am}^2 + N (1-h_{am}^2)}$$
, quando h_{am}^2 está associado à seleção de meios-irmãos (seleção entre famílias com sementes remanescentes).

Uma herdabilidade calculada usando médias ou totais de avaliações (medidas repetidas) por indivíduo pode ser convertida para o nível de indivíduo por:

(d)
$$h_{ai}^2 = \frac{h_{am}^2 [1 + (m-1)\rho]}{m}$$
, em que ρ é a repetibilidade e m o número de avaliações.

- (e) Uma herdabilidade de médias de clone $\{h_c^2 = Nh_{gi}^2/[1+(N-1) h_{gi}^2]\}$ pode ser convertida para herdabilidade individual no sentido amplo por
- $$h_{gi}^2 = \frac{h_c^2}{h_c^2 + N(1-h_c^2)}.$$

Outra formulação para h_c^2 é:

$$h_c^2 = \frac{Nh_{gi}^2}{1+(N-1) h_{gi}^2} = \frac{h_{gi}^2}{h_{gi}^2 + \frac{1-h_{gi}^2}{N}} = \frac{N}{N + \frac{1-h_{gi}^2}{h_{gi}^2}} = \frac{N}{N+\alpha}, \text{ em que:}$$

$$\alpha = \frac{1-h_{gi}^2}{h_{gi}^2}. \text{ Para } h_{gi}^2 = 0,50, \text{ tem-se } h_c^2 = \frac{N}{N+1}.$$

Outras formulações para h_c^2 e h_{am}^2 são $h_c^2 = 1-(1/F)$ e $h_{am}^2 = 1-(1/F)$, em que F é o valor do teste F de Snedecor da análise de variância, para a fonte de variação tratamentos (clones e progênies, respectivamente). No caso, h_{am}^2 é adequado à seleção entre famílias com sementes remanescentes.

14.10. Delineamento para a Maximização da Acurácia Seletiva em Testes Clonais

Os materiais a serem avaliados em testes clonais procedem das populações-base para seleção (testes de progênies) ou de seleção individual em áreas de ocorrência natural ou plantios comerciais. A primeira alternativa é preferível, pois permite maior acurácia neste primeiro estágio de seleção. Plantas individuais apresentam maior interação genótipo x ambiente e, portanto, os testes clonais devem ser instalados no maior número possível de ambientes. Dependendo da quantidade de material disponível, os testes clonais poderão ser realizados em dois estágios: (i) estágio inicial, visando à eliminação de clones com menor potencial produtivo; (ii) estágio final ou de recomendação de materiais para plantios comerciais.

O estágio inicial é recomendado quando se dispõe de um grande número de materiais genéticos para testes. Ele deve ser instalado em parcelas lineares e em um menor número de locais. Por outro lado, o estágio final exige maior rigor experimental, muitas vezes sendo necessária a utilização de parcelas quadradas, com bordaduras e com maior número de plantas. Todos esses fatores visam minimizar os efeitos de competição nas avaliações dos clones. Em ambos os estágios, os delineamentos a serem utilizados devem ser o de blocos casualizados (quando o número de materiais for pequeno) ou látice (quando o número de materiais for elevado).

Na Tabela 14.12 são apresentados os números adequados de indivíduos por clone em testes clonais, em função da herdabilidade individual no sentido

amplo.

Tabela 14.12. Número (nb) adequado de indivíduos por clone em testes clonais, em função da herdabilidade individual (h_g^2) no sentido amplo, para que sejam obtidas acurárias (r_{aa}) de 90% e 95%.

h_g^2	$r_{\text{aa}} = 90\%$	$r_{\text{aa}} = 95\%$	$r_{\text{aa}} \text{ para nb} = 100$
0,05	81	176	0,917
0,10	39	84	0,958
0,15	25	53	0,973
0,20	18	38	0,981
0,25	13	28	0,985
0,30	10	21	0,989
0,35	8	17	0,991
0,40	7	14	0,993
0,45	6	12	0,994
0,50	5	10	0,995
0,60	3	7	0,997
0,70	2	4	0,998
0,80	2	3	0,999
0,90	1	2	0,998

Fonte: Resende (1995).

Verifica-se pela Tabela 14.12, que 100 plantas por clone conduzem a acurárias superiores a 90%, independentemente da herdabilidade no sentido amplo. Assim, não se justifica empregar mais de 100 ramegos por clone. Entretanto, conhecendo-se a herdabilidade do caráter, pode-se optar por um número mais adequado. Por exemplo, com herdabilidade de 20%, consegue-se uma acurácia de 95% na seleção, empregando-se em torno de 40 plantas por clone e, acurácia de 90%, utilizando-se em torno de 18 ramegos por clone.

As expressões para o cômputo das acurárias são: $r_{\hat{g}g} = \left[\frac{b h_g^2}{1 + (b-1) h_g^2} \right]^{1/2}$ para o caso de uma planta por parcela e $r_{\hat{g}g} = \left[\frac{nb h_g^2}{1 + (n-1) (h_g^2 + c^2) + (b-1) nh_g^2} \right]^{1/2}$ para o caso de várias plantas por parcela .

Os valores apresentados na Tabela 14.12 resultam da 1^a expressão. Pela 2^a expressão, com $h_g^2 = 0,30$ e $c^2 = 0,10$, são necessários 15 indivíduos por clone para se obter acurácia de 90%.

14.11. Delineamentos para a Maximização da Acurácia Seletiva em Cruzamentos Fatoriais e Dialélicos

Para os delineamentos de cruzamentos fatoriais 4 x 4 ou dialélico parcial com cinco genitores, são mostrados, na Tabela 14.13, os valores de acurácia e valores de $N = nb$ para atingir determinada acurácia na seleção pelos valores genéticos aditivos associados aos vários indivíduos do experimento.

Tabela 14.13. Valores de acurácia e de $N = nb$, para atingir determinada acurácia na seleção por valores genéticos aditivos individuais, empregando os delineamentos de cruzamentos fatoriais 4 x 4 ou dialélico parcial com cinco genitores.

h_a^2	h_g^2	r_{aa} max.	N	% r_{aa} max	r_{aaN}
0,05	0,075	0,59	60	90	0,52
0,10	0,15	0,60	60	95	0,57
0,10	0,15	0,60	50	93	0,56
0,10	0,15	0,60	40	92	0,55
0,15	0,225	0,61	35	95	0,58
0,20	0,30	0,62	30	97	0,60
0,25	0,375	0,64	25	95	0,61
0,30	0,45	0,65	20	97	0,63
0,35	0,525	0,66	15	97	0,64
0,40	0,60	0,68	10	97	0,66

r_{aa} max. = acurácia máxima possível com infinitos indivíduos por família; r_{aaN} = acurácia atingida com N indivíduos por família; % r_{aa} max. = porcentagem de r_{aa} max. atingida com N indivíduos por família.

Estes valores foram obtidos com base nas expressões da acurácia apresentadas por Resende (1999b).

Verifica-se que entre 20 e 40 indivíduos por família de irmãos germanos são suficientes para atingir acima de 90% da acurácia máxima possível, para herdabilidades no sentido restrito entre 0,30 e 0,10, respectivamente.

14.12. Tamanho da Parcela

A questão do tamanho de parcelas experimentais em espécies perenes tem sido abordada por uma série de autores (Pearce, 1976; Cotterill & James, 1984; Pimentel Gomes, 1984; Resende, 1995; Bonnot, 1995, dentre outros).

Empregando-se a metodologia padrão baseada no coeficiente (b) de heterogeneidade do solo (Smith, 1938) (o qual é maior quanto maior for a heterogeneidade do solo), Bonnot (1995) indica os tamanhos de parcela: (i) 1 planta quando b for próximo de 0 ($b \leq 0,30$); (ii) pelo menos seis plantas quando b for próximo de 1 ($b \geq 0,70$). Como $0 \leq b \leq 1$, pode-se inferir que quando $0,3 < b < 0,7$, tamanhos de parcela entre 1 e 6, seriam recomendados.

Resende et al. (1996c) também empregaram a metodologia de Smith para determinar o tamanho de parcela em erva-mate, conforme descrito a seguir. Foram utilizadas 210 árvores de uma população plantada no município de Bocaiúva do Sul-PR, no espaçamento de 1,80 x 3,60 m, com mudas obtidas de sementes comerciais. Foram avaliados a altura, o diâmetro e o volume da copa, aos cinco anos de idade. Todas as plantas avaliadas eram equidistantes, tinham a mesma idade e estavam submetidas às mesmas condições ambientais, constituindo uma amostra aleatória de genótipos (não agrupados por família).

O delineamento experimental deste método de estimativa consistiu na subdivisão de uma amostra (número total de plantas avaliadas) de plantas em várias parcelas contíguas, constituídas por número (n) crescente de planta, a partir de $n = 1$. As mesmas plantas são utilizadas na demarcação imaginária das parcelas de diferentes tamanhos. No presente trabalho, foram avaliadas 15 linhas de 14 plantas (colunas), demarcadas no interior do talhão, gerando as 210 plantas mencionadas.

Para a geração dos diferentes tamanhos de parcela, foram considerados os gradientes ambientais, no sentido das linhas e das colunas, gerando dois grupos de dados para cada tamanho (n) de parcela maior que 1, conforme Tabela 14.14.

Tabela 14.14. Parcelas no sentido das linhas e colunas, para cada tamanho (n) de parcela maior que 1.

n	Linhas	Colunas
1	210	-
2	105	98
3	60	70
4	45	42
5	30	42
6	30	28
7	30	28
8	15	14
9	15	14
10	15	14

Para um mesmo tamanho de parcela, plantas diferentes devem ser usadas para gerar várias parcelas (parcelas de um mesmo tamanho devem ser independentes). Assim, para $n > 1$, os cálculos foram efetuados, separadamente, para cada um dos grupos (sentido das linhas ou das colunas referentes ao gradiente ambiental) e, posteriormente, os dois resultados foram combinados.

O procedimento estatístico para a estimativa de parâmetros genéticos baseou-se na mudança das variâncias das médias de parcelas como função da mudança dos tamanhos de parcelas. Assim, o mesmo conjunto total de plantas deve ser utilizado para o estabelecimento das parcelas de diferentes tamanhos. O modelo genético-estatístico associado à avaliação de uma planta i na parcela j equivale a:

$Y_{ij} = g_i + e_{ij}$, em que g_i é o efeito genotípico da planta i e e_{ij} é o efeito ambiental associado a planta i na parcela j ($i = 1, \dots, n; j = 1, \dots, p.$).

Tem-se que:

$Var(Y_{ij}) = \sigma_g^2 + \sigma_e^2$, em que σ_g^2 é a variância genética total (aditiva + não aditiva) e σ_e^2 é a variância ambiental.

Trabalhando-se em termos de médias de parcelas de diferentes tamanhos, tem-se:

$Var(Y_{\cdot j}) = \bar{\sigma}_{y(n)}^2 = \bar{\sigma}_g^2 + \bar{\sigma}_e^2$: variância entre médias de parcelas de tamanho n.

$$\bar{\sigma}_g^2 = \frac{\sigma_g^2}{n}$$

$\bar{\sigma}_e^2 = \frac{\sigma_e^2}{n^b}$, conforme demonstrado por Smith (1938), em que b é o coeficiente de heterogeneidade ambiental, variando de 0 a 1 à medida que o solo torna-se mais heterogêneo.

Por esta expressão vê-se que se $b=0$, $\bar{\sigma}_e^2 = \sigma_e^2$ e, se $b = 1$, $\bar{\sigma}_e^2 = \frac{\sigma_e^2}{n}$.

Considerando-se $b = 1$, a própria variância entre médias de parcela calculada pode ser utilizada para determinação do tamanho máximo da parcela.

Assim, a mudança em $\bar{\sigma}_y^2$ devida à variabilidade ambiental ($\bar{\sigma}_e^2$) é função logarítmica do tamanho de parcela. A mudança devida à variabilidade genética é função inversa do tamanho de parcela, pois considera-se que não existe correlação entre o valor genético e a localização das árvores no povoamento.

Assim:

$\bar{\sigma}_{y(n)}^2 = \frac{\sigma_g^2}{n} + \frac{\sigma_e^2}{n^b}$ ou, equivalentemente $n\bar{\sigma}_{y(n)}^2 = \sigma_g^2 + n^{(1-b)}\sigma_e^2$, em que:

$n\bar{\sigma}_{y(n)}^2$: quadrado médio entre médias de parcela de tamanho n, podendo ser estimado a partir dos dados experimentais.

$\sigma_g^2 + n^{(1-b)}\sigma_e^2$: esperança matemática do quadrado médio entre médias de parcela de tamanho n.

O estimador de $\bar{\sigma}_{y(n)}^2$ é dado por:

$\bar{\sigma}_{y(n)}^2 = \sum_{j=1}^p (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot \cdot})^2 \frac{1}{p-1}$, em que p é o número de parcelas associado a cada tamanho n.

As variâncias associadas às médias de parcelas de diferentes tamanhos (Tabela 14.15) revelaram que o tamanho ideal máximo ($b = 1$) de parcelas para

avaliação de materiais genéticos, ao nível de médias (procedências, progênies), é de seis plantas por parcela. Com o aumento do tamanho de parcela (n) de uma para seis plantas, houve reduções na variância fenotípica das médias de parcela. Com n maior que seis, houve aumento na variância fenotípica das médias de parcela, para os três caracteres avaliados. Isto decorre do fato de que, com o aumento do tamanho de parcelas (para $n > 6$), ocorre o aumento da variância ambiental (σ_e^2) em maior magnitude do que a redução em σ_e^2/n , devida ao aumento de n .

Tabela 14.15. Estimativas de variâncias associadas às médias de parcelas de diferentes tamanhos (n), em altura (AC), diâmetro (DC) e volume da copa (VC) de plantas de erva-mate, aos cinco anos de idade.

n	AC (m)	DC (m)	VC (m³)
1	0,4404	0,2767	2,5404 x 10 ⁻⁸
2	0,3169	0,1760	1,4537 x 10 ⁻⁸
3	0,2502	0,1640	1,2802 x 10 ⁻⁸
4	0,2254	0,1380	1,0651 x 10 ⁻⁸
5	0,1855	0,1312	8,7626 x 10 ⁻⁹
6	0,1504	0,1125	8,2927 x 10 ⁻⁹
7	0,1611	0,1177	9,4479 x 10 ⁻⁹
8	0,1572	0,1050	6,1834 x 10 ⁻⁹
9	0,1569	0,1071	6,4364 x 10 ⁻⁹
10	0,1713	0,1225	7,8194 x 10 ⁻⁹

Dessa forma, com base na metodologia de Smith, são recomendadas parcelas de uma a seis plantas, para experimentação com erva-mate.

Empregando, a metodologia do coeficiente de correlação intraclasse (relação variância entre parcelas/ (variância entre parcelas + variância dentro de parcelas)), Pimentel Gomes (1984) e Pimentel Gomes & Couto (1985), concluíram pelo uso de uma planta por parcela. Considerando a precisão experimental e a probabilidade de detecção de diferenças significativas entre médias de tratamentos, Cotterill & James (1984), Loo-Dinks & Tauer (1987), e Haapanen (1992) também concluíram pelo uso de uma planta por parcela.

No contexto da predição de valores genéticos, Resende (1995) realizou estudo avaliando a acurácia seletiva, para diferentes condições experimentais e níveis de herdabilidade: 5 diferentes valores da relação variância dentro de parcela/variância entre parcelas (5; 10; 20; 40; 80); 14 níveis de herdabilidade (5% a 90%); 30 diferentes combinações de tamanho (n) de parcela e número (b) de blocos (tamanhos de parcela variando de 1 a 10 plantas e número de blocos variando de 1 a 100).

Fixando um número total (nb) de indivíduos, verificou-se que parcelas com um indivíduo e várias repetições sempre conduzem à maior acurácia em relação a parcelas com vários indivíduos e menos repetições. Este fato ocorreu para todos os valores de herdabilidade e da relação $\sigma_\delta^2 / \sigma_e^2$ (variância dentro de parcelas/variância entre parcelas) testados.

As parcelas de uma planta podem acarretar o problema da perda de parcela, fato esse indesejável para a análise estatística, embora de menor efeito para a predição de valores genéticos. Por isso, a utilização de parcelas pequenas, porém com mais de uma planta, pode ser desejável. Outra alternativa, sugerida por Libby & Cockerham (1980), refere-se ao emprego de parcelas com várias plantas, mas não contíguas e sim com plantas distribuídas aleatoriamente no bloco. Segundo esses autores, este delineamento propicia eficiência comparável àquela obtida com parcelas de uma planta. Em cacaueiro, esta estratégia tem sido utilizada: considerando uma taxa de mortalidade de 10% durante os dois primeiros anos, Cilas (1995b) recomenda a utilização de quatro plantas por tratamento, aleatorizados dentro de cada bloco.

Em resumo, parcelas de uma a seis plantas devem ser empregadas na experimentação com espécies perenes, embora alguns autores recomendem fortemente (White, 1996) o uso de apenas uma planta por parcela. O uso de apenas uma planta por parcela permite a avaliação de 100 a 150 entradas por bloco, enquanto mantém o bloco em tamanho aceitável de cerca de 0,1 ha conforme recomendado para as condições tropicais (Matheson, 1989).

O único problema da perda de parcelas de uma planta, em termos de avaliação genética, refere-se à perda de representatividade de alguns genitores em determinados níveis dos efeitos fixos (blocos), o que pode tornar as comparações genéticas viciadas, se não for adotado o procedimento BLUP.

Com relação ao fator competição entre tratamentos (entre clones e entre progênies), o uso de parcelas de uma planta é muito vantajoso. Isto porque a planta de um tratamento é circundada por plantas de 8 tratamentos distintos, de forma que os efeitos da competição intergenotípica sobre uma planta tendem a serem cancelados, uma vez que tal planta central tem a chance de ser circundada por plantas de tratamentos mais e também menos agressivos. Este resultado é similar àquele obtido quando se usam parcelas (com bordaduras) de 9 plantas, avaliando-se apenas a planta central. Porém, o uso de parcelas de uma planta é mais indicado, pois permite avaliar 8 vezes o número de plantas, na mesma área, permitindo, assim, a avaliação de um maior número de tratamentos. Os efeitos da competição intragenotípica podem ser considerados utilizando plantios piloto com apenas um genótipo (clone) ou seja, por meio de plantios monoclonais para a confirmação da seleção.

Por outro lado, com parcela de várias plantas, cada tratamento é circundado por praticamente apenas dois tratamentos, o que pode permitir a manifestação dos efeitos (na média geral do tratamento) de competição, se poucas repetições forem utilizadas.

14.13. Tamanho da População Experimental, Número de Famílias e Número de Repetições

O tamanho da população experimental (número total de indivíduos a serem avaliados), deve ser determinado considerando a maximização do ganho genético com seleção, o qual depende simultaneamente da maximização da acurácia seletiva e da intensidade de seleção. Resende (1995) realizou estudo neste sentido, obtendo as seguintes conclusões:

- Em termos de ganho genético (que depende simultaneamente da acurácia e da intensidade de seleção), constatou-se que, independentemente da herdabilidade do caráter, não é vantajoso avaliar mais de 20.000 indivíduos nos experimentos. Com coeficientes de herdabilidade maiores que 20%, cerca de 10.000 indivíduos conduzem a eficiências satisfatórias; pouco ganho (menos de 5%) é conseguido com a duplicação do número de indivíduos para 20.000.
- Com herdabilidade baixa e um número total de indivíduos fixo, é compensador avaliar menos famílias (não menos que 50) e maior número de indivíduos por família.

Para a estimativa de parâmetros genéticos, White (1996) também recomenda um número mínimo de 50 genitores, sendo 100 o número preferido. Uma discussão detalhada sobre o número de genitores a serem usados nos programas de melhoramento de espécies perenes é apresentada por Resende (1999a).

A determinação do número de repetições é realizada dividindo-se o número total de indivíduos por família, necessários à maximização da acurácia seletiva ou à precisa estimativa de parâmetros genéticos, pelo tamanho de parcela escolhido.

14.14. Delineamentos Experimentais

Um delineamento experimental adequado deve obedecer aos princípios fundamentais da experimentação: repetição, casualização e controle local. A importância do número de repetições é capital, significando que, com baixo número de repetições, até a casualização é prejudicada ou comprometida. Como controle local deve ser enfatizada a homogeneidade dentro de estratos ou blocos, sendo, em princípio, recomendados os delineamentos em blocos casualizados e látice. A casualização e a repetição é que propiciam uma comparação não viciada dos tratamentos, ao passo que o controle local e a repetição permitem reduzir o erro experimental médio. Um erro experimental menor permite inferir como significativa uma diferença real pequena entre médias de tratamentos.

O delineamento em quadrado latino, provavelmente, é o que propicia melhor controle local, visto que permite controlar a heterogeneidade ambiental em

duas direções, no sentido das linhas e das colunas. Entretanto, tal delineamento não tem sido recomendado para os trabalhos de melhoramento (Ramalho et al., 2000) ou na experimentação em geral (Pimentel Gomes, 1987), devido à restrição do número de repetições ter que ser igual ao número de tratamentos ou progênies. Dessa forma, não há relatos de sua utilização no melhoramento. No entanto, com o advento da utilização de parcelas de uma planta no melhoramento de plantas perenes, tal delineamento passa a ter grande potencial de utilização. Como se utiliza em torno de 60 plantas por progénie (60 repetições de uma planta), quadrados latinos de 60 x 60 com 60 progênies poderiam ser perfeitamente utilizados, em associação com o procedimento BLUP. No caso, os dados seriam corrigidos para dois gradientes ambientais (linhas e colunas), pelo método BLUP ou do índice multifeitos. Segundo Panse & Sukhatme (1963), quando existem tendências simultâneas de variações em fertilidade em duas direções em ângulos retos (que equivale a uma tendência diagonal em fertilidade), é provável que o quadrado latino seja mais eficiente que o delineamento em blocos. O delineamento em quadrado latino é também recomendado quando não se conhece a priori os gradientes de fertilidade.

Os delineamentos de blocos incompletos (látice, por exemplo) são especialmente indicados na situação de grande número de tratamentos e alta variabilidade ambiental (quantificada pelo b de Smith, por exemplo) na área experimental. No melhoramento de espécies florestais, o delineamento de blocos casualizados tem sido o mais utilizado na América do Norte (Fu et al., 1998), os blocos incompletos são os mais utilizados na Austrália, África do Sul e Ásia (Williams & Matheson, 1994) e, no Brasil, ambos os tipos de delineamentos são empregados.

A eficiência relativa entre os delineamentos experimentais depende, sobretudo, do nível de variação ambiental espacial na área experimental. Empregando um modelo geoestatístico espacial, o qual permite a especificação de vários níveis de variação ambiental, Fu et al. (1998) concluíram pela superioridade dos delineamentos de blocos incompletos (látice e alfa) em um grande número de situações, em termos de eficiência estatística para a estimação de médias de tratamentos.

Outra classe de delineamentos que tem sido utilizada de maneira crescente nos últimos anos são os blocos aumentados de Federer (Federer, 1958, 1998; Wolfinger et al., 1997), os quais, por construção são desbalanceados. Os delineamentos de blocos aumentados, blocos incompletos balanceados e parcialmente balanceados não são ortogonais. Neste caso, o uso destes delineamentos para seleção conduz, via análise intrablocos, a médias de tratamentos imprecisas, mesmo quando a sobrevivência for 100% (Resende, 1999b). Não significa isto, que estes delineamentos não devem ser usados, mas que sejam usados em associação com o procedimento BLUP, o qual ajusta as médias para as estimativas BLUE dos efeitos fixos ou efeitos ambientais identificáveis.

Entretanto, os delineamentos em blocos baseiam-se na premissa de conhecimento a priori da heterogeneidade da área experimental de forma que seja

possível alocar todas as parcelas (tratamentos) em blocos homogêneos (Lotode, 1971). Caso esta heterogeneidade não seja conhecida a priori, a delimitação dos blocos tornar-se-á arbitrária, fato que poderá implicar forte heterogeneidade dentro de blocos. Com base no exposto, uma alternativa é a alocação aleatória de parcelas de uma planta no campo experimental e posterior controle da heterogeneidade ambiental, empregando métodos tais como o da análise de covariância associando uma covariável à variável estudada (Método de Papadakis: Papadakis, 1984), ou das variáveis regionalizadas ou espaciais (Métodos Geostatísticos, conforme Lecoustre & Reffye, 1986).

O ajustamento a posteriori para os gradientes ambientais em testes de progênies apresenta um potencial para um aumento significativo da eficiência na estimação de parâmetros genéticos e seleção. Neste contexto, o uso de parcelas de uma planta completamente aleatorizadas (delineamento inteiramente ao acaso) tem sido novamente comum (Lotode & Lachenaud, 1988). Entretanto, Gilmour (2000) adverte para o fato de que a blocagem a posteriori não deve ser baseada apenas na significância estatística de contrastes arbitrários. O experimentador necessita identificar as causas físicas e ambientais que levaram à determinado tipo de blocagem. É importante relatar que o próprio uso do delineamento em quadrado latino pode reduzir a necessidade do uso de técnicas (como a análise espacial) de ajustamento a posteriori. Mais detalhes sobre a eficiência da blocagem em experimentos de campo são apresentados no tópico 19.11.

Os vários delineamentos são, em geral, adequados para a estimação de parâmetros genéticos e para a seleção de genitores com base no comportamento de suas progênies. Para estes objetivos, o delineamento experimental de blocos casualizados, com famílias de meios-irmãos e uma planta por parcela, é especialmente recomendado (White, 1996) como o mais apropriado também para a estimação de correlações genéticas e estudos de interação genótipo x ambiente.

Para a seleção de indivíduos da população experimental, duas possibilidades existem: (i) seleção dos próprios indivíduos avaliados (em parcelas com uma ou poucas plantas) no experimento; (ii) seleção em parcelas de progênies fora do experimento, sem delineamento, usando as informações das médias de progénie no experimento e os valores individuais nas parcelas fora do experimento.

Na opção (i), a seleção deve ser feita pelo índice multifeitos (IME) ou BLUP (Capítulo 5). Na opção (ii), o índice para a predição do valor genético é dado por:

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

$$b_1 = \frac{(1/4) \sigma_a^2}{\sigma_y^2}, \text{ para meios-irmãos.}$$

$$b_2 = \frac{(3/4) \sigma_a^2}{\sigma_\delta^2 + \sigma_e^2}.$$

\bar{Y}_i = média da família na parcela fora do experimento.

σ_δ^2 : variância dentro de parcelas.

σ_e^2 : variância entre parcelas.

Considerando o mesmo número de indivíduos (60) por progênies no experimento e fora do experimento tem-se:

- a intensidade de seleção é a mesma.
- a homogeneidade para a comparação na situação (ii) é maior do que em (i), visto que em (ii) o estrato homogêneo tem 60 plantas e em (i) o estrato tem cerca de 100 a 150 plantas (considerando a avaliação de 100 a 150 progênies em cada bloco). Isto pode ser afirmado uma vez que cada planta da progénie (parcelas de uma planta) é ajustada pela média do bloco.
- Maior b_2 em (ii) do que em (i), exatamente devido a exposto em (b).
- Maior b_1 em (i) do que em (ii) (Resende, 1991).

Com $N = 60$, (d) pode ser dado por:

$$\frac{b_{1(ii)}}{b_{1(i)}} = \frac{\frac{(1/4)}{1+(N-1)(1/4)}}{N} = \frac{0,25}{0,2625} = 0,95$$

A superioridade de (i) ou (ii) advém, então, do balanço entre (c) e (d), o qual depende, sobretudo, da heterogeneidade ambiental da área experimental. Provavelmente as duas abordagens conduzirão a resultados similares.

Um outro delineamento proposto (Fasoulas, 1981) em literatura é o de anéis hexagonais (*honeycomb*). Neste delineamento, a competição é evitada ao máximo possível. No caso, uma planta central é comparada com várias (seis) plantas vizinhas, de forma que os efeitos de competição intergenotípica cancelam. A disposição das plantas em um hexágono regular permite que todas as plantas fiquem eqüidistantes e a comparação de uma planta com as suas vizinhas imediatas permite melhor avaliação de seu valor genotípico. Segundo a filosofia deste delineamento uma planta só é selecionada se a ela for superior à média de suas seis vizinhas. Assim, tal delineamento pode ser efetivo em controlar os efeitos ambientais.

Segundo Gallais (1989), com este delineamento, é possível selecionar, por meio de um índice combinando a performance individual e a performance média das plantas vizinhas. Wricke & Weber (1986) afirmam que não existem ainda resultados experimentais que comprovem ou não a superioridade deste delineamento sobre os demais. No Brasil, tal delineamento foi aplicado à seleção de clones de guaraná, por Escobar (1986).

Uma situação experimental peculiar refere-se à avaliação de progênies pertencentes a várias populações, no delineamento em blocos ao acaso. A aleatorização de todas as progênies (independente de populações) nos blocos é um procedimento que permite tanto uma comparação entre populações quan-

to entre progênies dentro de populações. Entretanto, o termo de erro para comparação entre progênies/populações pode ser subestimado para algumas e superestimado para outras progênies, pois tal erro pode diferir dentro das populações. Um arranjo experimental especial denominado “blocos de famílias compactas” foi proposto por Panse & Sukhatme (1963), o qual permite reduzir o erro na comparação entre progênies dentro de populações. Tal arranjo é análogo mas não idêntico ao arranjo em parcelas subdivididas. Trata-se de um arranjo hierárquico no qual as populações são sorteadas nas parcelas e as progênies dentro de populações são sorteadas nas subparcelas. A diferença para o arranjo em parcelas subdivididas é que, no caso dos blocos de famílias compactas, os tratamentos das subparcelas (as progênies dentro de populações) são diferentes de uma parcela (população) para outra.

O arranjo experimental de blocos de famílias compactas permite realizar análises de variância individuais para cada população e, portanto, considerar um termo de erro (variância residual) específico para cada população. Quando existir homogeneidade de variância do erro dentro de populações, um termo de erro comum pode ser utilizado para testar as fontes de variação de progênies dentro de cada população e, também, a fonte de variação progênies/população em geral.

O uso apenas da fonte de variação global progênies/população é recomendado quando se dispõe de um reduzido número de progênies dentro de cada população, fato que não permite análises dentro de cada população, para fins de estimativa de parâmetros genéticos. Neste caso, o procedimento mencionado aumenta consideravelmente o número de graus de liberdade da fonte de variação progênies/populações, permitindo a estimativa de uma herdabilidade única, válida para todas as populações. A análise das progênies de todas as populações como sendo uma única população é reprovável, pois tem gerado superestimativas da herdabilidade, uma vez que tal procedimento distribui a variação entre populações como se fosse variação genética aditiva entre progênies, o que não é verdadeiro.

Na prática, o uso deste arranjo é especialmente recomendado quando as populações apresentam diferenças substanciais em taxas de crescimento. Os blocos de famílias compactas permitem minimizar os efeitos de competição e também contornar o problema da heterogeneidade de variâncias entre populações. Maiores detalhes da análise de experimentos no delineamento de blocos de famílias compactas são apresentados nos tópicos 7.9, 9.8.3, 9.8.5 e no capítulo 19.

Finalmente, é importante destacar que os testes de comparações de médias não apresentam qualquer relevância no contexto do delineamento de experimentos genéticos com espécies perenes. Uma vez que, por definição, os efeitos genéticos são considerados variáveis aleatórias, não existe qualquer significado para os testes de comparação de médias (Newman-Keuls, Duncan, Scheffé, Tukey, Bonferroni, Fisher, etc.), cujos métodos são derivados na suposição de efeitos fixos de tratamentos, isto é, constantes fixas e não variáveis ale-

tórias. Detalhes adicionais sobre a avaliação de tratamentos de efeitos fixos são apresentados no capítulo 13.

14.15. Número Adequado de Locais de Experimentação para Um Número Total Fixo de Indivíduos a Serem Avaliados

O número adequado de locais de experimentação para um número total fixo de indivíduos depende da herdabilidade do caráter e da correlação genética intraclasse através dos locais (\hat{r}_{gg}). Fixando-se em $N\ell$ o número total de indivíduos por acesso, em que N refere-se ao número de indivíduos por local, tem-se os seguintes estimadores (considerando as herdabilidades de mesma magnitude em todos os locais) para as variâncias fenotípicas ao nível de médias de acesso, considerando a avaliação dos $N\ell$ indivíduos em um só ambiente ou em vários ambientes (Resende, 1998) são:

$$\hat{\sigma}_{y\ell}^2 = \left(\hat{h}^2 + \frac{1-\hat{h}^2}{N\ell} \right) \hat{\sigma}_{yi}^2$$

$$\hat{\sigma}_{ym}^2 = \left[\hat{r}_{gg} \hat{h}^2 + \frac{(1-\hat{r}_{gg})}{\ell} \hat{h}^2 + \frac{(1-\hat{h}^2)}{N\ell} \right] \hat{\sigma}_{yi}^2$$

em que:

$\hat{\sigma}_{y\ell}^2$ e $\hat{\sigma}_{ym}^2$: estimativas das variâncias fenotípicas ao nível de médias de acesso para o local ℓ e na média dos ℓ locais, respectivamente.

$\hat{\sigma}_{yi}^2$: estimativa da variância fenotípica individual dentro de locais.

$\hat{h}^2 = \frac{\hat{\sigma}_g^2 + \hat{\sigma}_{ge}^2}{\hat{\sigma}_{yi}^2}$: estimativa da herdabilidade individual dentro de locais, podendo ser no sentido restrito (teste de progênie) ou amplo (teste clonal).

$\hat{\sigma}_g^2$: estimativa da variância genética livre da interação genótipo x ambiente.

$\hat{\sigma}_{ge}^2$: estimativa da variância da interação genótipo x ambiente.

A eficiência (E) da seleção baseada em vários locais em relação à seleção baseada em um só local é dada por:

$$E = \frac{G_{s_2}}{G_{s_1}} = \left[\frac{\hat{\sigma}_g^2 + \hat{\sigma}_{ge}^2 + \hat{\sigma}_{\delta}^2/n\ell}{\hat{\sigma}_g^2 + \frac{\hat{\sigma}_{ge}^2}{\ell} + \frac{\hat{\sigma}_{\delta}^2}{n\ell}} \right]^{1/2} = \left[\frac{\hat{h}^2 + (1-\hat{h}^2)/n\ell}{\hat{r}_{gg} \hat{h}^2 + \frac{(1-\hat{r}_{gg})}{\ell} \hat{h}^2 + (1-\hat{h}^2)/n\ell} \right]^{1/2} = \left[\frac{1+(n\ell-1) \hat{h}^2}{1+(n-1) \hat{h}^2 + n (\ell-1) \hat{r}_{gg} \hat{h}^2} \right]^{1/2}$$

em que, $\hat{\sigma}_{\delta}^2$ refere-se a variação residual dentro de locais.

Na Tabela 14.16 são apresentados valores de eficiência para alguns valores de n , ℓ e r_{gg} para $\hat{h}^2 = 0,20$. Para $\hat{h}^2 = 0, 20, 30$ indivíduos por família conduzem a acurárias da ordem de 90% da máxima possível para a seleção de indivíduos para propagação por sementes ou de clones para propagação vegetativa (Resende, 1995).

Tabela 14.16. Eficiência da experimentação em vários locais (ℓ) em relação a um só local, considerando um número total ($N\ell$) fixo de indivíduos avaliados, para várias magnitudes de correlação genética (r_{gg}) intraclasse através dos locais, número (N) de indivíduos por acesso e herdabilidade (h^2) de 20% (genética, podendo ser no sentido restrito ou amplo).

h^2	$N\ell$	ℓ	N	r_{gg}	E	Incremen-to (%)
0,20	30	2	15,0	0,30	1,20	20
		3	10,0	0,30	1,30	10
		4	7,5	0,30	1,38	8
		5	6,0	0,30	1,40	2
0,20	30	2	15,0	0,50	1,13	13
		3	10,0	0,50	1,19	6
		4	7,5	0,50	1,22	3
		5	6,0	0,50	1,24	2
0,20	30	2	15,0	0,70	1,07	7
		3	10,0	0,70	1,10	3
		4	7,5	0,70	1,12	2
		5	6,0	0,70	1,13	1

Verifica-se, com base na Tabela 14.16, que para um número total fixo de indivíduos avaliados, torna-se vantajosa (ganho de 6%) a utilização de 4, 3 e 2 locais para correlações genéticas entre locais de magnitudes da ordem de 0,30; 0,50 e 0,70, respectivamente.

14.16. Estimativas de Parâmetros e Controle Genético de Caracteres de Importância Econômica em Algumas Espécies Perenes

Conforme verificado nos tópicos anteriores, as estimativas de parâmetros genéticos, além de suas importâncias na inferência sobre o controle genético dos diferentes caracteres e comparação de métodos de seleção, são essenciais

nos procedimentos de predição de valores genéticos e, portanto, na seleção propriamente dita.

Na predição de valores genéticos, os componentes de variância são considerados como conhecidos, mas, a rigor, para a avaliação genética com dados desbalanceados ou não, necessita-se apenas de estimativas fidedignas dos parâmetros das populações base (originais) não selecionadas e não endogâmicas. Para a obtenção destas estimativas o melhorista tem duas opções: (i) a partir de seus próprios dados experimentais; (ii) a partir da literatura. Cabe ao melhorista, em função das circunstâncias, discernir qual das duas opções propicia estimativas mais confiáveis. O uso de estimativas de parâmetros, obtidas de literatura, é uma forma de uso de informação a priori.

Exceto para as espécies florestais, estimativas dos parâmetros genéticos úteis (h_a^2 , h_g^2 e ρ ao nível de indivíduo) para a predição de valores genéticos, são raramente reportadas para as espécies perenes, e a maioria dos estudos refere-se a herdabilidades no sentido amplo ao nível de médias de clones ou de cruzamentos, as quais não são úteis para a predição de valores genéticos individuais.

Estimativas obtidas de literatura, para a herdabilidade individual no sentido restrito (h_a^2) para alguns caracteres de crescimento em *Eucalyptus grandis* e *Eucalyptus urophylla* são apresentadas nas Tabelas 14.17 e 14.18, respectivamente.

Tabela 14.17. Estimativas da herdabilidade individual no sentido restrito para caracteres de crescimento, densidade da madeira e resistência ao cancro em *Eucalyptus grandis*.

Altura	Diâmetro	Volumen	Densidade	Resistência	Nº. de Progê-nies	Nº. de Estimativas	Idade (anos)	País	Referência
0,11	0,08	0,10	-	-	51	1	1	África do Sul	Van Wyk (1976)
0,45	0,30	0,53	-	-	71	1	8	África do Sul	Van Wyk (1990)
0,39	0,39	0,31	-	-	529	1	5	EUA	Reddy & Rockwood (1989)
0,25	0,25	0,27	-	-	63 a 72	4	5	África do Sul	Verry et al. (1997)
-	-	0,38	-	-	163	21	2,5	Brasil	Duda et al. (1997)
0,23	0,23	-	-	-	282	7	3	Brasil	Resende & Higa (1992)
0,19	0,22	0,20	-	-	25	2	3 e 5	Brasil	Resende et al. (1994b)
0,22	0,16	-	-	-	39	1	3	Brasil	Resende & Higa (1994a)
0,18	0,16	0,17	-	-	33	1	3	Brasil	Resende & Bertolucci (1995)
0,32	0,18	-	-	-	64	1	3	Brasil	Oda et al. (1989)
0,45	0,47	-	-	-	-	2	2	Brasil	Borges (1980)
0,26	0,15	-	-	-	64	3	2	Brasil	Kageyama (1980)
0,28	0,13	0,10	0,69	-	64	3	7	Brasil	Moraes (1987)
-	-	-	0,82	-	-	1	3,5	Brasil	Kageyama et al. (1983a)
0,27	0,25	0,31	-	-	81	2	2,5	Brasil	Kikuti (1988)
0,35	0,32	0,29	-	-	180	56	5	Brasil	Higa et al. (1993)
0,08	0,10	0,10	-	-	101	3	2 a 4	Brasil	Duratex (1993)
-	-	-	-	0,70	124	3	2	Brasil	Borges & Brune (1981)

Continua...

Tabela 14.17. Continuação.

Altura	Diâmetro	Volume	Densidade	Resistência	Nº. de Progê-nies	Nº. de Estimativas	Idade (anos)	País	Referência
0,07	0,09	0,12	-	-	178	3	3,2	África do Sul	Snedden et al. (2000)
0,19	0,16	0,20	-	-	178	3	5,5	África do Sul	Snedden et al. (2000)
0,22	0,22	0,20	-	-	180	5	2,7	Uruguai	Balmelli (2000)
0,19	0,20	0,18	-	-	180	4	4,7	Uruguai	Balmelli (2000)
Média Ponderada									
0,28	0,27	0,27	0,76		0,70				

Tabela 14.18. Estimativas da herdabilidade individual no sentido restrito para caracteres de crescimento e densidade da madeira em *Eucalyptus urophylla*.

Altura	Diâmetro	Volume	Densidade	Nº. de Progê-nies	Nº. de Estimativas	Idade (anos)	País	Referência	
0,36	0,24	-	-	67	4	3	Brasil	Pinto Júnior (1984)	
0,39	0,24	0,25	-	66 a 109	4	7	Brasil	Mori et al. (1988)	
0,50	0,43	0,44	-	146	2	8	Brasil	Santos et al. (1990)	
-	0,23	-	-	45	1	3	Brasil	Rezende & Bertolucci (1993)	
-	0,35	-	-	45	1	4	Brasil	Rezende & Bertolucci (1993)	
-	0,35	-	-	45	1	5	Brasil	Rezende & Bertolucci (1993)	
0,15	0,17	-	-	33	1	5	Brasil	Freitas et al. (1997)	
0,32	0,24	-	0,71	90	4	1,2,3 e 4	China	Wei & Borralho (1995)	
-	0,15	-	-	45	1	1	Brasil	Resende (1999b)	
-	0,18	-	-	45	1	2	Brasil	Resende (1999b)	
-	0,21	-	-	45	1	3	Brasil	Resende (1999b)	
-	0,23	-	-	45	1	4	Brasil	Resende (1999b)	
-	0,25	-	-	45	1	5	Brasil	Resende (1999b)	
-	0,25	-	-	45	1	6	Brasil	Resende (1999b)	
-	0,25	-	-	45	1	7	Brasil	Resende (1999b)	
Média Ponderada									
0,36	0,26	0,31	0,71						

Resultados médios de estimativas da herdabilidade individual no sentido restrito para várias espécies florestais foram relatados por Cornellius (1994) e são apresentados na Tabela 14.19.

Tabela 14.19. Estimativas médias* de herdabilidade individual no sentido restrito para várias espécies florestais, baseadas em 528 estimativas.

	Altura	Diâmetro	Volume	Densidade
Média	0,28	0,23	0,21	0,50
Mediana	0,25	0,19	0,18	0,48

* Adaptado de Cornellius (1994).

Na Tabela 14.20 são apresentadas estimativas da herdabilidade individual no sentido restrito para o caráter volume em *Eucalyptus grandis* e *E. urophylla* na Aracruz Celulose, conforme Rezende & Resende (2000).

Tabela 14.20. Estimativas* da herdabilidade individual (h_a^2) no sentido restrito para o caráter volume em *Eucalyptus grandis* e *E. urophylla* na Aracruz Celulose.

Espécie	Local	h_a^2	c^2
<i>E. grandis</i>	Aracruz	0,27	0,02
	Aracruz	0,62	-
	Aracruz	0,28	-
	Aracruz	0,14	-
	Aracruz	0,20	-
	Aracruz	0,18	-
	Aracruz	0,21	-
	Aracruz	0,32	-
<i>E. urophylla</i>	Aracruz	0,43	0,02
	São Mateus	0,19	0,07
	Sul da Bahia	0,25	0,04
	Aracruz	0,69	0,04
	São Mateus	0,34	0,08
	Sul da Bahia	0,70	0,001
Média		0,34	0,04

* c^2 - correlação fenotípica entre indivíduos devida ao ambiente comum da parcela.

Na Tabela 14.21 são apresentadas estimativas da herdabilidade individual no sentido amplo para o caráter volume em *Eucalyptus grandis*, *E. urophylla* e híbridos, conforme Rezende & Resende (2000).

Tabela 14.21. Estimativas* da herdabilidade individual (h_g^2) no sentido amplo para o caráter volume, obtidas a partir da avaliação de clones de *Eucalyptus grandis*, *E. urophylla* e híbridos (resultados estratificados de acordo com o número de clones na amostra).

	h_g^2	c^2	σ_y^2	Nº. de clones	Nº. de estimativas
Média	0,31	0,11	0,006726	9 a 121	81
Mediana	0,30	0,10	0,001577	9 a 121	81
Média	0,30	0,11	0,007830	24 a 121	67
Mediana	0,30	0,10	0,001664	24 a 121	67
Média	0,33	0,11	0,001442	9 a 16	14
Mediana	0,33	0,09	0,001231	9 a 16	14

* c^2 e σ_y^2 - correlação fenotípica entre indivíduos decorrente do ambiente comum da parcela e variância fenotípica ao nível de indivíduo, respectivamente.

Os resultados médios da herdabilidade no sentido restrito para os caracteres de crescimento em *Eucalyptus grandis* apresentaram magnitudes próximas aos resultados médios obtidos para o *E. urophylla* (Tabelas 14.17 e 14.18), equivalendo a, aproximadamente, 0,30, valor este que pode ser considerado de magnitude moderada e que é superior aos valores médios (Tabela 14.19) descritos para espécies florestais (principalmente coníferas) por Cornellius (1994).

Os resultados médios da herdabilidade no sentido amplo para o caráter volume em *Eucalyptus* também revelaram uma magnitude em torno de 0,30. Levando em conta que estes valores médios possam representar parâmetros de uma única população, ter-se-iam, a princípio, valores coincidentes das herdabilidades nos sentidos amplo e restrito (concordando com os resultados de Kikuti, 1988) e, consequentemente, uma variância não aditiva (ou de dominância, negligenciando a epistasia) nula. Entretanto, os valores da herdabilidade no sentido restrito foram obtidos, tratando as populações de *Eucalyptus* como alógamas. Empregando marcadores enzimáticos, relatos de literatura (Eldridge et al., 1993), apresentam para o *E. grandis* em populações naturais, uma taxa de autofecundação de 15%. Como a maioria (Higa et al., 1993) das estimativas obtidas advieram de populações naturais recém-introduzidas, torna-se necessário considerar os efeitos do sistema reprodutivo misto nas estimativas da herdabilidade.

Para uma taxa de autofecundação de 15%, sob sistema reprodutivo misto, os coeficientes de parentesco médio entre indivíduos de progêniens de polinização aberta equivalem a 0,33 e 0,30, para caracteres sem e com dominância, respectivamente (Resende et al., 1995c). Como na obtenção das estimativas da herdabilidade no sentido restrito foi considerado o coeficiente de parentesco de 0,25, a estimativa média deve ser corrigida (multiplicada) pelos fatores 0,25/0,33 e 0,25/0,30, para o caso de ausência e presença de dominância, respectivamente, tornando-se razoável considerar h^2 em torno de 0,25. Este valor é mais próximo do valor médio para espécies florestais, de acordo com Cornellius (1994).

Sendo $h_a^2 = 0,25$ e $h_g^2 = 0,30$, a herdabilidade dos efeitos de dominância (d^2) equivale a cerca de 0,05, ou seja, ao redor de 20% de h_a^2 (significando que $\sigma_d^2 / \sigma_a^2 = 0,20$, aproximadamente), para uma população hipotética pouco melhorada ou com nível intermediário de melhoramento. Este resultado, se confirmado, revela a presença de dominância (conforme esperado, o que explica a ocorrência de depressão endogâmica) e a baixa magnitude dos efeitos de dominância em relação aos efeitos aditivos, fato que poderia explicar a baixa heterose verificada em ambientes favoráveis, relatada por Martin (1989). A baixa magnitude dos efeitos de dominância (herdabilidade no sentido amplo pouco superior à herdabilidade no sentido restrito) também foi mostrada por Snedden et al. (2000).

Nas Tabelas 14.22, 14.23 e 14.24 são apresentadas estimativas de parâmetros genéticos em nível individual para algumas espécies perenes.

Tabela 14.22. Estimativas do parâmetro genético herdabilidade individual no

sentido restrito (h_a^2) em algumas espécies perenes.

Pinus caribaea* var. *hondurensis

Caráter	Idade (anos)	h_a^2	Referência
Volume	6,5	0,11	Ledig & Whitmore (1981)
	5,0	0,18	Resende et al. (1995b)
	5,0	0,30	Sampaio et al. (2000)
	3,0	0,17	Camcore (1997)
	5,0	0,19	Camcore (1997)
	8,0	0,16	Camcore (1997)
Altura	6,5	0,10	Ledig & Whitmore (1981)
	2,0	0,36	Kageyama et al. (1983b)
	8,0	0,29	Dean et al. (1986); Cotterill & Dean (1990)
	5,0-7,0	0,20	Woolaston et al. (1990); Cotterill & Dean (1990)
	5,0	0,06	Sampaio et al. (2000)
	9,5	0,22	Resende (1999b)
	6	0,24	Powell & Nikles (1996)
	6	0,15	Dungey et al. (2000)
	10	0,21	Dungey et al. (2000)
Diâmetro	6,5	0,19	Ledig & Whitmore (1981)
	8,0	0,48	Dean et al. (1986); Cotterill & Dean (1990)
	5,0-7,0	0,30	Woolaston et al. (1990); Cotterill & Dean (1990)
	5,0	0,20	Sampaio et al. (2000)
	9,5	0,13	Resende (1999b)
	6	0,10	Powell & Nikles (1996)
	6	0,13	Dungey et al. (2000)
	10	0,20	Dungey et al. (2000)
Média	6,5	0,20	

***Acácia-negra* (*Acacia mearnsii*)**

Caráter	h_a^2	Referência
Diâmetro	0,20; 0,11; 0,34; 0,30	Wright (1976); Resende (1999a); Resende et al. (1998a)
Altura	0,08; 0,30	Resende et al. (1999a)
Teor de tanino	0,60; 0,28	Wright (1976); Resende (1999a)
Relação tanantes/não tanantes	0,21	Resende (1999a)
Resistência à gomose	0,05	Resende (1999a)
Resistência ao serrador	0,04	Resende (1999a)
Espessura da casca	0,20	Wright (1976)

Erva-mate (*Ilex paraguariensis*)

Caráter	h_a^2	Referência
Massa foliar	0,19; 0,27; 0,18	Resende et al. (1995a) Resende et al. (2000b)
Volume da copa	0,07	Resende et al. (1995a)
Enraizamento de estacas	0,06 – 0,25	Resende et al. (1995a)

Seringueira (*Hevea brasiliensis*)

Caráter	h_a^2*	Nº. de estimativas	Referência
Produção de borracha	0,25 0,09 0,18 0,50	12 1 1 1	Resende et al. (1996d)* Costa et al. (2000) Costa et al. (2000) Costa et al. (2000)

* : média dos dados de literatura.

Dendê (*Elaeis guineensis*)

Caráter	h_a^2	Referência
Número de cachos	0,35	Soh (1994)
Peso do cacho	0,20	Soh (1994)
Óleo/cacho	0,15	Soh (1994)
Frutos/cacho	0,15	Soh (1994)
Frutos/cacho	0,18	Hartley (1977)

Maçã (*Malus domestica*)

Caráter	h_a^2	Referência
Tamanho do fruto	0,34	Durel et al. (1998)
Textura	0,34	Durel et al. (1998)
Sabor	0,39	Durel et al. (1998)
Conteúdo de suco	0,39	Durel et al. (1998)
Atratividade	0,38	Durel et al. (1998)
Circunferência do tronco	0,51	Durel et al. (1998)
Resistência ao mildeio	0,68	Durel et al. (1998)
Aspecto da fruta	0,79	Currie et al.(2000)
Conicidade	0,38	Currie et al.(2000)
Squareness	0,35	Currie et al.(2000)

Pêra (*Pyrus pyrifolia*)

Caráter	h_a^2	Referência
Peso de frutos	0,73	Abe et al. (1995)
Firmeza	0,29	Abe et al. (1995)
Conteúdo de Sólidos Solúveis	0,42	Abe et al. (1995)
pH do suco da fruta	0,64	Abe et al. (1995)
Tempo para amadurecimento	0,95	Abe et al. (1993)
Forma do fruto: extensão	0,42	White et al.(2000)
Forma do fruto: profundidade	0,44	White et al.(2000)

Café (*Coffea canephora*)

Caráter	h_a^2	Referência
Peso de sementes	0,15 a 0,91	Leroy et al. (1994)
Peso de sementes	0,73	Montagnon et al. (1998)
Conteúdo de cafeína	0,80	Montagnon et al. (1998)
Conteúdo de gordura	0,74	Montagnon et al. (1998)
Conteúdo de sacarose	0,11	Montagnon et al. (1998)

Café (*Coffea arabica*)

Caráter	h_a^2	Referência
Altura da planta	0,37	Cilas et al. (1998a)
Diâmetro do caule	0,06	Resende et al.(2001c)
Peso de frutos	0,26	Cilas et al. (1998a)

Cacau (*Theobroma cacao*)

Caráter	h_a^2	Referência
Podridão parda	0,11	Cilas et al. (1998b)
Número de frutos por planta	0,10	Resende & Dias (2000)
Peso de sementes / planta	0,21	Dias & Resende (2001b e c)
Número de frutos sadios	0,22	Dias & Resende (2001b e c)
Peso de sementes / fruto	0,22	Dias & Resende (2001b e c)

Acerola (*Malpighia emarginata*)

Caráter	h_a^2	Referência
Peso anual de frutos	0,31	Paiva et al. (2002)

Pupunha (*Bactris gasipaes*)

Caráter	h_a^2	Referência
Altura da planta	0,18	Farias Neto & Resende (2000)
Diâmetro da planta	0,10	Farias Neto & Resende (2000)
Tamanho do palmito	0,42	Farias Neto & Resende (2000)
Peso do palmito	0,06	Farias Neto & Resende (2000)

Cupuaçu (*Theobroma grandiflorum*)

Caráter	h_a^2	Referência
Peso médio de frutos	0,41	Souza et al. (2000)
Número de frutos / planta	0,33	Souza et al. (2000)

Tabela 14.23. Estimativas do parâmetro genético herdabilidade individual no sentido amplo (h_g^2) em algumas espécies perenes.

Pinus caribaea* var. *hondurensis

Caráter	h_g^2	Referência
Diâmetro	0,39	Resende (1999b)
Altura	0,92	Resende (1999b)

Seringueira (*Hevea brasiliensis*)

Caráter	$h_g^2 *$	Nº. de estimativas	Referência
Produção de borracha	0,30	4	Resende et al. (1996d)

* = média dos dados de literatura.

Erva-mate (*Ilex paraguariensis*)

Caráter	$h_g^2 *$	Nº. de estimativas	Referência
Enraizamento de estacas	0,27	3	Resende et al. (1995a)
Massa foliar	0,28	1	Resende et al. (2000b)

* = estimativa média.

Dendê (*Elaeis guineensis*)

Caráter	h_g^2	Referência
Número de cachos	0,70	Baudouim & Durandgasselin (1991)
Peso médio do cacho	0,50	Baudouim & Durandgasselin (1991)
Produção de cachos	0,40	Baudouim & Durandgasselin (1991)
Frutos/Cacho	0,30	Baudouim & Durandgasselin (1991)
Mesocarpo/Fruto	0,60	Baudouim & Durandgasselin (1991)
Óleo/Mesocarpo	0,50	Baudouim & Durandgasselin (1991)
Óleo/Cacho	0,40	Baudouim & Durandgasselin (1991)
Produção de óleo	0,40	Baudouim & Durandgasselin (1991)

Cacau (*Theobroma cacao*)

Caráter	h_g^2	Referência
Altura da planta	0,62 e 0,77	Valois & Nascimento (1986)
Diâmetro do caule	0,24	Valois & Nascimento (1986)
Número de frutos por planta	0,38	Resende & Dias (2000)
Peso de sementes / planta	0,28	Dias & Resende (2001b e c)
Número de frutos sadios	0,26	Dias & Resende (2001b e c)
Peso de sementes / fruto	0,24	Dias & Resende (2001b e c)
Podridão parda	0,13	Cilas et al. (1998b)

Guaraná (*Paullinia cupana*)

Caráter	h_g^2	Referência
Número de folhas	0,33	Escobar (1986)
Comprimento do maior ramo	0,33	Escobar (1986)
Número de ramos	0,31	Escobar (1986)

Café (*Coffea arabica*)

Caráter	h_g^2	Referência
Altura da planta	0,49	Cilas et al. (1998a)
Diâmetro do caule	0,12	Resende et al.(2001c)
Peso de frutos	0,40	Cilas et al. (1998a)
Peso de frutos	0,17	Walyaro e Vossen (1979)

Caju (*Anacardium occidentale*)

Caráter	h_g^2	Referência
Altura da planta	0,40	Crisóstomo et al. (1999)

Tabela 14.24. Estimativas do parâmetro genético repetibilidade individual (ρ) em algumas espécies perenes.

Espécie	Caráter	ρ	Nº. Estima-tivas	Referência
<i>Eucalyptus urophylla</i>	Circunferência do tronco	0,92	21	Resende (1999b)
	Incremento anual na circunferência	0,69	1	Resende (1999b)
Erva-mate	Massa foliar	0,70	3	Resende et al. (1995a)
Seringueira	Produção de borracha	0,80	4	Resende et al. (1996d)
Cacau-eiro	Número de frutos sadios	0,50	1	Dias & Souza (1993)
	Peso de sementes úmidas por planta	0,53	1	Dias & Souza (1993)
	Peso de sementes úmidas por frutos	0,67	1	Dias & Souza (1993)
	Número de frutos por planta	0,57		Resende & Dias (2000)
Coqueiro	Número de frutos por planta	0,45	1	Siqueira (1982)
Cupuaçu	Comprimento do fruto	0,82	1	Costa et al. (1997)
	Circunferência do fruto	0,50	1	Costa et al. (1997)
	Espessura da casca	0,93	1	Costa et al. (1997)
	Peso do fruto	0,77	1	Costa et al. (1997)
	Número de sementes	0,55	1	Costa et al. (1997)
	Porcentagem de casca	0,63	1	Costa et al. (1997)
	Porcentagem de polpa	0,70	1	Fonseca et al. (1990)
	Porcentagem de sementes	0,72	1	Fonseca et al. (1990)
	Porcentagem de acidez da polpa	0,63	1	Fonseca et al. (1990)
	Porcentagem de casca	0,50	1	Fonseca et al. (1990)
	Peso da polpa	0,61	1	Fonseca et al. (1990)
	Relação brix/acidez	0,60	1	Fonseca et al. (1990)
	Peso da casca	0,58	1	Fonseca et al. (1990)
	Largura do fruto	0,56	1	Fonseca et al. (1990)
Guaranazeiro	Peso total	0,56	1	Fonseca et al. (1990)
	Tamanho da inflorescência	0,77	1	Valois et al. (1979)
	Número de botões	0,74	1	Valois et al. (1979)
	Número de frutos	0,66	1	Valois et al. (1979)
Acerola	Número de sementes	0,74	1	Valois et al. (1979)
	Peso anual de frutos	0,70	1	Paiva et al. (2001a)
Média		0,66		

De maneira geral, a classificação dos coeficientes de repetibilidade quanto à magnitude pode ser realizada comparando o ganho fenotípico permanente a ser obtido considerando uma medição (G_1) com aquele a ser atingido, levando em conta m medições (G_m) de onde se obtém a relação $G_1/G_m = \{[1+(m-1) \rho]/m\}^{1/2}$. Sendo o padrão $m = 2$, para G_m as seguintes classificações para a repetibilidade são:

- repetibilidade alta: $G_1/G_m \geq 0,90 \Rightarrow r \geq 0,60$
- repetibilidade média: $0,80 < G_1/G_m < 0,90 \Rightarrow 0,30 < r < 0,60$
- repetibilidade baixa: $G_1/G_m \leq 0,80 \Rightarrow r \leq 0,30$

Com base nesta classificação, pode-se inferir que a maioria dos caracteres quantitativos de espécies vegetais perenes apresenta repetibilidade alta. O valor médio encontrado foi de 0,66, indicando que mesmo com h_a^2 baixa (em torno de 0,20 para os referidos caracteres em erva-mate, seringueira, *Pinus* e *Eucalyptus*) a repetibilidade é alta. Considerando que h_g^2 situa-se ao redor de 0,30 para seringueira e para *Eucalyptus*, verifica-se que a fração $p^2 = \sigma_p^2 / \sigma_y^2$ provavelmente seja alta (em torno de 0,35 a 0,50) para espécies vegetais perenes, revelando um forte componente ambiental permanente. Em animais domésticos, as repetibilidades verificadas (Willis, 1991) não são tão altas, citando-se os valores de 0,45 para produção de leite ($h_a^2 = 0,30$) em bovinos e 0,15 para tamanho e peso da leitegada em suínos, de forma que o componente p^2 parece ser bem mais baixo. De fato, a superioridade de p^2 para espécies vegetais perenes em relação àquelas verificadas para os animais é esperada em função de os animais estarem menos sujeitos a efeitos ambientais permanentes devido a suas capacidades de movimento.

Para espécies como o coqueiro, que apresenta flutuações bianuais de produção (Shrikhande, 1957), as baixas estimativas de repetibilidade obtidas podem ser atribuídas ao tipo de análise empregada. Nestes casos, o método da análise de variância subestima a repetibilidade (Vasconcellos et al., 1985).

14.17. Obtenção de Estimativas Médias de Parâmetros Genéticos por meio de Ponderações

A obtenção de estimativas agrupadas (ou médias) de componentes de variância e de parâmetros genéticos pode ser realizada por uma ponderação baseada no inverso da variância de cada estimativa isolada (Cunningham et al., 1977). Para a herdabilidade, por exemplo, tem-se:

$$\hat{h}^2 = \sum_{i=1}^n \hat{w}_i \hat{h}_i^2 : \text{estimativa ponderada da herdabilidade, em que:}$$

$\hat{w}_i = (1/\hat{\sigma}_i^2) / (\sum_{i=1}^n (1/\hat{\sigma}_i^2))$: coeficiente de ponderação da estimativa da herdabilidade proveniente do experimento ou população ou local i.

n: número de estimativas usadas na ponderação.

\hat{h}_i^2 : estimativa da herdabilidade no experimento i.

$\hat{\sigma}_i^2$: estimativa da variância da herdabilidade no experimento i.

O desvio padrão da estimativa ponderada de h^2 é dado por

$$\hat{s}(\hat{h}^2) = \left(\sum_{i=1}^n \hat{w}_i^2 \hat{\sigma}_i^2 \right)^{1/2}.$$

Procedimento análogo pode ser adotado para a ponderação de estimativas de componentes de variância. Os estimadores apresentados são muito úteis em compilações de resultados experimentais e/ou de literatura.

14.18. Classificação das Magnitudes da Herdabilidade e da Acurácia Seletiva

A acurácia está intimamente ligada à herdabilidade do caráter. Uma classificação da herdabilidade e acurácia em termos de magnitude e suas associações é apresentada a seguir, conforme Resende (1997).

Herdabilidade individual	Acurácia para seleção individual	Acurácia máxima possível para a seleção combinada usando também a média de família	Classificação das magnitudes da herdabilidade individual	Classificação das magnitudes da acurácia para seleção de indivíduos
0,01	0,10	0,51	Baixa	Baixa
0,10	0,32	0,55	$0,01 \leq h_a^2 \leq 0,15$	$0,10 \leq r_{aa} \leq 0,40$
0,15	0,39	0,58		
0,20	0,45	0,61	Média ou	Média ou
0,30	0,55	0,66	Moderada	moderada
0,40	0,63	0,71	$0,15 < h_a^2 < 0,50$	$0,40 < r_{aa} < 0,70$
0,50	0,71	0,76	Alta	Alta
0,60	0,77	0,80	$h_a^2 \geq 0,50$	$r_{aa} \geq 0,70$
0,70	0,84	0,85		
0,80	0,89	0,90		
0,90	0,95	0,95		

Verifica-se que, com $h_a^2 > 0,50$, praticamente não existe vantagem no uso da informação de família e a seleção com base apenas na informação de indivíduo já propicia uma acurácia alta ($> 0,70$). Mesmo para caracteres de baixa herdabilidade, o uso da informação de parentes (maior número de informações)

permite elevar a acurácia seletiva da classe de baixa para moderada. Este fato ressalta a importância de trabalhar com métodos mais elaborados de seleção.

14.19. Análise Espacial via REML/BLUP Individual

14.19.1. Introdução

A análise tradicional de experimentos de campo parte do princípio de que todas as observações tomadas em posições adjacentes (em plantas ou parcelas vizinhas) são não correlacionadas. Desta forma, a matriz de covariância residual ou dos erros é modelada como $R = I\sigma_e^2$, ou seja, os erros são assumidos como independentes. Também, a posição dos tratamentos no campo, ou seja, a sua distribuição espacial é ignorada. Segundo Steel & Torrie (1980), a casualização concorre para a neutralização dos efeitos da correlação espacial e, portanto, para a geração de uma análise de variância fidedigna.

Entretanto, conforme Grondona et al. (1996), embora a teoria da casualização enfatize a neutralização da correlação espacial, tal neutralização é mais eficiente quando se usam modelos espaciais. Também as formas de controle local baseadas em blocagem podem ser ineficientes para tratar de problemas de gradientes ambientais e mesmo os blocos incompletos não permitem uma avaliação completa dos efeitos espaciais. Além disso, a blocagem é realizada antes da implantação dos experimentos, de forma que se percebe muitas vezes (principalmente em espécies perenes), por ocasião da coleta dos dados experimentais, a presença de manchas ou gradientes ambientais dentro dos experimentos, os quais não foram considerados adequadamente pelos blocos delineados a priori. Nesta situação, somente as técnicas de análise espacial permitem contornar a questão e propiciar uma seleção acurada, através de blocagem a posteriori ou através da flexibilização da matriz R baseados nos próprios dados experimentais.

A autocorrelação espacial advinda de observações com dependência espacial compromete as suposições de homogeneidade dentro de blocos e de ausência de interação genótipos x blocos, suposições estas essenciais para obter eficiência no delineamento experimental e boa capacidade de teste. Assim, a constatação de ocorrência de autocorrelação espacial e a consideração de uma estrutura de erros correlacionados com $R = \Sigma$, em que Σ refere-se a uma matriz não diagonal, podem permitir a obtenção de estimativas e previsões mais precisas, aumentando assim, a eficiência da análise estatística.

A consideração deste tipo de estrutura ($R = \Sigma$) é também relevante na análise de medidas repetidas tomadas sobre um mesmo indivíduo em plantas perenes, sendo uma opção aos modelos de repetibilidade, multivariado e de regressão aleatória. A autocorrelação espacial ou serial é positiva e aumenta com a diminuição da distância entre as observações. Dessa forma, modelos que consideram a correlação entre medidas repetidas como sendo a mesma para quaisquer pares ou combinações de idades ou safras podem não ser os mais

adequados. Assim sendo, os modelos de análise espacial podem ser vantajosos também para a análise de medidas repetidas, ou seja, para análise de dados com dependência temporal.

O presente tópico discorre sobre a aplicação de métodos geoestatísticos e de análise de séries temporais na análise de experimentos de campo delineados no contexto do melhoramento de plantas perenes. Ênfase é dada à análise de dados com dependência espacial, mas extensões à análise de dados com dependência temporal são triviais.

O objetivo final do melhoramento em benefício da produção vegetal refere-se à maximização da expressão fenotípica do caráter de interesse no campo do produtor rural. Uma vez que a expressão fenotípica é função do genótipo, do ambiente e da interação genótipo x ambiente, torna-se relevante o estudo destes três fatores. Neste contexto, a predição dos valores genéticos é fundamental, mas o estudo dos fatores ambientais é de igual relevância.

Visando à separação dos efeitos genéticos e ambientais, são utilizadas técnicas de genética quantitativa aplicadas sobre dados obtidos de experimentação a campo. Uma tentativa de refinamento no estudo dos genótipos diz respeito à utilização de técnicas de biologia molecular, notadamente a análise QTL's (locos controladores de características quantitativas) marcados. Um refinamento no estudo dos efeitos ambientais refere-se ao emprego de técnicas de análise espacial (via modelos geoestatísticos e de séries temporais), as quais permitem uma análise mais detalhada da variabilidade espacial dos solos, permitindo melhor estratificação ambiental (identificação de estratos mais homogêneos e de gradientes ambientais) e comparação mais efetiva dos genótipos através do uso de uma estrutura de erros espacialmente dependentes e correlacionados.

14.19.2. Métodos de análise espacial e áreas de aplicação em genética

A variabilidade ou heterogeneidade espacial associada à fertilidade e estrutura do solo, umidade, interceptação de luz e outros fatores ambientais contribuem para o aumento da variação residual. Assim, é importante controlar, por delineamento ou por análise, a variação residual (Grondona et al., 1996).

Além dos delineamentos experimentais, outras formas de controle local e aumento da precisão experimental referem-se à análise de covariância ou ajuste de covariável e aos métodos de análise de vizinhança. Os métodos de análise de vizinhança fundamentam-se no princípio de vizinhança e no ajuste de uma covariável associada a observações de parcelas vizinhas, sendo que os dois principais métodos são: (i) médias móveis (MA), baseado na média das parcelas vizinhas; (ii) Papadakis, baseado na média dos erros das parcelas vizinhas. Segundo Pearce (1998), ao usar o método de Papadakis não é necessário usar a blocagem simultaneamente.

Outros procedimentos destinados ao controle da correlação positiva entre observações obtidas de plantas vizinhas são: (i) método das variáveis regionalizadas ou procedimentos geoestatísticos (Cressie, 1993; Martinez, 1994); (ii) métodos de análise de séries temporais, via modelos AR, MA e ARIMA em duas dimensões (Martin, 1990; Cullis & Gleeson, 1991; Gilmour et al., 1997; Gilmour et al., 1998; Cullis et al. 1998); (iii) modelo linear de campo aleatório (Zimmerman & Harville, 1991); (iv) ajuste de superfícies de resposta polinomiais (Federer, 1998).

O método de análise de superfícies de tendências baseia-se no ajuste de superfícies de resposta polinomiais ou de *splines* e foi recomendado por Federer (1998). Entretanto, segundo Gilmour (2000), tais modelos raramente se justificam na prática. Verbyla et al. (1999) reportam que a tendência global pode ser acomodada mediante o ajuste de *splines* cúbicas. O ajuste de tais *splines* pode ser realizado por meio do software ASREML (Gilmour et al., 2000).

O modelo linear de campo aleatório foi desenvolvido por Zimmermann & Harville (1991) e considera diretamente a heterogeneidade espacial por meio da inclusão dos efeitos de tendência e das correlações entre os erros, modelando de forma direta a matriz de covariância dos erros. Esta modelagem é similar à empregada nos métodos geoestatísticos para fazer previsões espaciais, pois consideram as observações como sendo uma realização parcial de um campo aleatório. O método considera dependência espacial em todas as direções e preserva quaisquer esquemas de blocagem. Tal modelo procura uma estimativa da função geral de covariância, a qual participa diretamente dos procedimentos de estimação e previsão.

Esta função é da forma $R = \sigma^2 [f(h)]$, em que $f(h)$ é uma função da distância entre as observações nas posições físicas s e $(s + h)$, em que cada posição é definida por um par de coordenadas dado pela linha e coluna na grade experimental.

Tem-se $Cov (e_s, e_{(s+h)}) = Cov (h) = \sigma^2$ se $h = 0$ e

$Cov (e_s, e_{(s+h)}) = \sigma^2 [f(h)]$, se $h > 0$. A determinação da função geral que descreve a covariância espacial pode ser realizada pela abordagem geoestatística, ajustando uma função contínua ao variograma amostral, conforme detalhado na seqüência.

No método das variáveis regionalizadas, utilizam-se os procedimentos geoestatísticos para avaliar o padrão de variabilidade espacial na área experimental, obtendo-se a matriz de semivariâncias ajustadas, a qual é utilizada como ponderadora no sistema de equações de quadrados mínimos generalizados (Cressie, 1993; Martinez, 1994) ou modelo misto. Martinez (1994) empregou tal método usando semivariâncias ajustadas pelo modelo esférico.

As principais ferramentas geoestatísticas utilizadas na análise espacial são (Valente, 1989; Ribeiro Júnior, 1995; Landim, 1998; Duarte, 2000; Soares, 2000; Vieira, 2000):

(i) Cômputo das semivariâncias

A semivariância é definida como $S(h) = (1/2) \operatorname{Var}[e_{(s+h)} - e_{(s)}]$, ou seja, equivale à metade da variância de diferenças entre observações (entre erros, no caso) separadas por uma distância h . Valores baixos de $S(h)$ indicam menor variabilidade, ou seja, maior similaridade.

Dentre os vários estimadores de semivariâncias, o mais utilizado é o estimador clássico de Matheron, baseado no método dos momentos e dado por:

$$\hat{S}(h) = \frac{1}{2N(h)} \sum_{N(h)} [\hat{e}_{(s+h)} - \hat{e}_{(s)}]^2, \text{ em que } N(h) \text{ é o número de diferenças tomadas à distância } h.$$

(ii) Construção de semivariogramas ou variogramas

O semivariograma ou variograma representa uma função de semivariâncias em relação às suas respectivas distâncias e é utilizado para a estimação da estrutura de variabilidade espacial como função da distância entre as observações. O variograma amostral é construído plotando-se os valores calculados de semivariância no eixo das ordenadas contra as respectivas distâncias no eixo das abscissas, ou seja, plotando-se os pontos $[h, \hat{S}(h)]$.

Se os valores de $\hat{S}(h)$ distribuem-se aleatoriamente em função de h , pode-se inferir sobre independência entre observações ou resíduos. Se $\hat{S}(h)$ crescem com o aumento de h , pode-se inferir sobre dependência espacial entre observações ou resíduos, revelada pela menor variabilidade a menores distâncias.

(iii) Realização de Krigagem

O semivariograma amostral não é suficiente para descrever a variabilidade em toda a área, uma vez que se baseia em estimativas de semivariâncias apenas para algumas distâncias. Uma interpolação por Krigagem permite descrever a variabilidade para toda a área, tendo por base a estimativa da estrutura de dependência entre observações vizinhas e o ajuste de uma função contínua de semivariograma a partir do semivariograma amostral discreto.

(iv) Obtenção da função que descreve a covariância espacial

A partir do ajuste de uma função contínua ao variograma amostral, obtém-se a função de covariância espacial $\operatorname{Cov}(e_s, e_{(s+h)})$ pela relação $\operatorname{Cov}(e_s, e_{(s+h)}) = \operatorname{Cov}(h) = \sigma^2 - S(h)$, em que $\sigma^2 = \operatorname{Cov}(h=0)$. As funções de variograma mais comumente utilizadas são os modelos esférico, exponencial e gaussiano, os quais correspondem às seguintes funções de autocovariância:

Modelo esférico

$$Cov(e_s, e_{(s+h)}) = \begin{cases} \sigma^2 \left[1 - \frac{3}{2} \left(\frac{h}{\alpha} \right) + \frac{1}{2} \left(\frac{h}{\alpha} \right)^3 \right], & \text{se } h < \alpha \\ 0, & \text{se } h \geq \alpha \end{cases}$$

Modelo exponencial

$$Cov(e_s, e_{(s+h)}) = \sigma^2 \exp\left(\frac{-h}{\alpha_e}\right), \text{ em que } \alpha_e = \frac{1}{3} \alpha$$

Modelo gaussiano

$$Cov(e_s, e_{(s+h)}) = \sigma^2 \exp\left(\frac{-h^2}{\alpha^2}\right), \text{ em que } \alpha \text{ refere-se ao alcance da cor-}$$

relação espacial ou à distância a partir da qual a semivariância estabiliza.

Os métodos de análise de séries temporais foram utilizados inicialmente por Gleeson & Cullis (1987) que consideraram os erros por meio de um processo auto-regressivo integrado de médias móveis (ARIMA (p, q, d) em uma única direção. Este modelo foi considerado ineficiente por considerar os efeitos espaciais em uma só dimensão. Martin (1990) e Cullis & Gleeson (1991) estenderam tal modelo a duas dimensões: linhas e colunas. Tal modelo estendido é da forma ARIMA (p₁, d₁, q₁) x ARIMA (p₂, d₂, q₂).

Esses modelos são denominados modelos com erros nas variáveis e consideram um efeito de tendência (ξ) mais um erro η independente. A diferença entre os vários métodos reside na modelagem e estimação do efeito de tendência. Nos métodos mencionados, a tendência é modelada por um processo ARIMA, o qual se limita a diferenças de primeira ordem ou diferenças entre observações adjacentes. Alguns métodos somente incluem ξ e ignoram η .

Em ensaios com culturas agrícolas, na ausência de conhecimento da correta estrutura de correlação, Gilmour et al. (1997) sugerem a modelagem de ξ como um processo auto-regressivo separável de primeira ordem (AR1 x AR1), ignorando η . Este processo auto-regressivo bidimensional de primeira ordem tem-se mostrado eficiente em uma gama de situações (Apitolaza et al., 2000; Gilmour, 2000; Costa e Silva et al., 2001), algumas com a inclusão de η .

Os métodos de análise espacial têm, recentemente, sido aplicados a várias subáreas da genética e do melhoramento. No Brasil, trabalho relevante foi realizado por Duarte (2000), aplicando e comprovando a eficiência da análise espacial via métodos geoestatísticos no contexto de melhoramento de plantas anuais, por meio de implementação via o Programa SAS (Statistical Analysis System). Na Austrália, Gilmour & Thompson (1998) e Gilmour et al. (2000) implementaram o software ASREML, destinado à análise espacial de experimen-

tos em melhoramento de plantas anuais e perenes, permitindo a análise ao nível de indivíduos e o emprego de vários procedimentos de análise espacial com ênfase nos métodos de análise de séries temporais.

No contexto da análise de medidas repetidas com dependência temporal, o relevante trabalho de Apiolaza et al. (2000) demonstrou a utilidade dos modelos auto-regressivos com variâncias heterogêneas. Os procedimentos de análise espacial têm sido aplicados também no estudo de dados moleculares, associados à genética de populações separadas espacialmente (Degen & Scholz, 1998; Lecorre et al., 1998) e também na análise de QTL's considerando os efeitos de posição no cromossomo, revelando a grande utilidade da abordagem espacial para a genética e o melhoramento.

Tendo em vista os bons resultados práticos propiciados pelos procedimentos de análise de séries temporais, este tópico enfatiza mais esta técnica. Ênfase também é dada à análise espacial ao nível de indivíduos via abordagens de modelos lineares mistos (REML/BLUP) visando à estimativa de componentes de variância e à predição de valores genéticos no melhoramento de plantas perenes.

14.19.3. Modelos ARIMA na análise espacial de experimentos de campo

Dentre os modelos paramétricos de análise de séries temporais, o método de Box e Jenkins tem sido um dos mais usados (Box & Jenkins, 1970). Este método consiste em ajustar modelos auto-regressivos-integrados-médias móveis, ARIMA (p, d, q) a um conjunto de dados. No caso, p , d e q referem-se à ordem do processo auto-regressivo, ao número de diferenças ou de diferenciações para tornar a série estacionária e à ordem do processo médias móveis, respectivamente. Dependendo dos valores de p , q e d , os modelos ARIMA são equivalentes aos modelos: médias móveis (MA), quando $p = d = 0$; auto-regressivos (AR) quando $d = q = 0$; auto-regressivo-médias móveis (ARMA) quando $d = 0$. Assim, tem-se:

$$\begin{aligned} \text{ARIMA } (p, 0, 0) &= \text{AR}(p); \\ \text{ARIMA } (0, 0, q) &= \text{MA } (q); \\ \text{ARIMA } (p, 0, q) &= \text{ARMA } (p, q). \end{aligned}$$

Então, os modelos AR, MA e ARMA usam a série original, supondo estacionariedade.

A modelagem da estrutura de erro como um processo estocástico estacionário foi considerada inicialmente por Martin (1982; 1986). Posteriormente, Gleeson & Cullis (1987) propuseram modelar a tendência nos experimentos de campo como um efeito aleatório representado por um processo ARIMA de baixa ordem, em apenas uma dimensão. Cullis & Gleeson (1991) estenderam para duas dimensões esta abordagem, a qual tem sido usada com grande eficiência (Grondona et al., 1996; Gilmour et al., 1997; Gilmour, 2000). A seguir será apresentada esta abordagem.

Considerando os dados de parcela indexados na ordem de campo, tem-se o seguinte modelo segundo Gleeson & Cullis (1987):

$y = D\tau + \xi + \eta$, em que:

y : vetor de dimensão n contendo as observações.

τ : vetor de dimensão t , contendo os efeitos de tratamentos.

D : matriz de incidência para τ , de dimensões ($n \times t$).

ξ : vetor de dimensão n , representando os efeitos aleatórios de tendência.

η : vetor de dimensão n , representando os efeitos localizados, considerados como desvios independentes $N(0, \sigma^2)$.

Considerando a série temporal ξ_t e assumindo que os elementos ξ_t ($t = 1, \dots, n$) de ξ do modelo podem ser representados por um processo ARIMA (p, d, q), tem-se na notação de Box & Jenkins (1970):

$\phi(B)\Delta^d \xi_t = \theta(B) a_t$, em que:

B : operador de translação para o passado, $B\xi_t = \xi_{t-1}$.

d : nível de diferenciação necessário para fazer $\Delta^d \xi_t$ estacionário, $\Delta = 1 - B$ e $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$.

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$.

a_t : ruídos brancos assumidos como desvios independentes $N(0, \sigma_a^2)$.

Assumindo que os elementos η_t ($t = 1, \dots, n$) de η são ruídos brancos adicionais, Gleeson & Cullis (1987) mostraram que $\Delta^d (\xi_t + \eta_t)$ é um processo ARMA (p, Q) com Q sendo o maior dentre ($p+d$) e q . Tais autores relataram também que uma pequena subclasse de modelos ARIMA ($p, d, 0$) com $p = 0$ ou 1 e $d = 1$ ou 2 geralmente propiciam adequado ajustamento.

Cullis & Gleeson (1991) consideraram um experimento com r linhas e c colunas e denotaram por $W = (w_{ij})$ a matriz dos dados indexados na ordem de campo. Fazendo-se $w = \text{vec}(W)$, o modelo para w é dado por:

$w = D\tau + \xi + \eta$, em que:

τ : vetor de dimensão t contendo os efeitos de tratamentos.

ξ : vetor de efeitos aleatórios de tendência, com dimensão $N = rc$.

η : vetor de erros localizados, com dimensão $N = rc$.

D : matriz de incidência para τ com dimensões ($N \times t$).

O vetor ξ pode ser representado por um processo ARIMA (p_2, d_2, q^*) ao longo das linhas. Denotando $\varepsilon_i = \xi_i + \eta_i$ como um erro na parcela (ou planta) i , tem-se na notação de Box e Jenkins para modelos multiplicativos:

$\Phi(B^r)\Delta_r^{d_2} \varepsilon_i = \Theta(B^r) \alpha_i$, em que:

$\Delta_r = 1 - B^r$, sendo que B é o operador de translação para o passado.

$\Phi(B^r)$ e $\Theta(B^r)$: polinômios em B^r de graus p_2 e q_2 , respectivamente, no qual $q_2 = \max(q^*, p_2 + d_2)$.

Este modelo liga os componentes do erro ε_i , r parcelas adiante em w . E se ε_i são correlacionados também nas colunas, os componentes a_i, a_{i-1}, \dots serão correlacionados. Para considerar a existência de correlações nas colunas, deve-se introduzir um segundo modelo dado por:

$$\phi(B)\Delta^{d_1}a_i = \theta(B) a_i, \text{ em que:}$$

a_i : ruído branco

$\phi(B)$ e $\theta(B)$: polinômios em B de graus p_1 e q_1 , respectivamente, em que $\Delta = 1 - B$.

Substituindo o modelo para colunas, no modelo para linhas, tem-se o seguinte modelo geral multiplicativo.

$$\phi_{p_1}(B) \Phi_{p_2}(B^r) \Delta^{d_1} \Delta_r^{d_2} \varepsilon_i = \theta_{q_1}(B) \Theta_{q_2}(B^r) a_i.$$

Este processo pertence à subclasse de processos látice os quais são denominados **separáveis**. Se $\rho(g, 0)$ e $\rho(0, h)$ são as correlações lag-g e lag-h de processos unidimensionais ao longo de colunas e linhas, respectivamente, então a **separabilidade** implica

$$\rho(g, h) = \rho(g, 0) \rho(0, h).$$

Martin (1979) mostrou que para a realização de um processo separável em um látice retangular, a matriz de variância equivale ao produto de Kronecker das matrizes de variância de processos unidimensionais. Assim, para os dados diferenciados, tem-se:

$$y = (\Delta_c \otimes \Delta_r) w = \Delta w = \Delta D \tau + e = X \tau + e, \text{ em que:}$$

$$Var(e) = \sigma^2 H.$$

$$H = \sum_2(\gamma_2) \otimes \sum_1(\gamma_1).$$

γ_1 e γ_2 : vetores $(p_1 + q_1)$ e $(p_2 + q_2)$ de parâmetros dos processos Arma nas linhas e colunas, ou seja, parâmetros da matriz de covariância.

$$X = (\Delta_c \otimes \Delta_r) D.$$

Δ : matriz $n_1 \times r$ que promove a diferenciação para a ordem d_1 .

Δ_c : matriz $n_2 \times c$ que promove a diferenciação para a ordem d_2 .

$$n_1 = r - d_1.$$

$$n_2 = c - d_2.$$

Este modelo geral, desenvolvido por Cullis & Gleeson (1991) é, essencialmente, equivalente ao desenvolvido por Martin (1990).

No contexto dos experimentos de melhoramento, o seguinte modelo linear geral pode ser especificado:

$$y = D\tau + Z\beta + S + \eta, \text{ em que:}$$

y : vetor de dados, ordenados como colunas e linhas dentro de colunas.

τ : vetor representando os efeitos de tratamentos ou de genitores.

β : vetor representando a variação espacial em larga escala ou tendência global (efeitos de blocos, tendência polinomial).

S : representa a variação espacial em pequena escala (dentro de blocos) ou tendência local, modelada como um vetor aleatório com média zero e variância espacialmente dependente.

η : vetor de erros independentes e identicamente distribuídos.

Em termos de processos ARIMA separáveis, S pode ser escrito como:

$$\phi_{p_1}(B) \Phi_{p_2}(B^r) \Delta_r^{d_1} \Delta_r^{d_2} S_i = \theta_{q_1}(B) \Theta_{q_2}(B^r) a_i$$

Usando um modelo deste tipo, Grondona et al. (1996) concluíram que, dentre 19 modelos avaliados em 35 ensaios, aqueles que consideram a correlação espacial em duas dimensões são os mais eficientes, destacando-se o modelo ARIMA (1, 0, 0) x ARIMA (1, 0, 0), ou seja, o modelo AR(1) x AR(1). O modelo AR(1) x AR(1) mostrou-se o mais eficiente em 21 dos 35 ensaios.

No modelo AR(1), em analogia ao modelo auto-regressivo $\xi = \phi \xi_{i-1} + a_i$ para séries temporais e sendo $\phi = \rho_1$ a estimativa inicial de ϕ , tem-se que o processo AR(1) padrão para os erros é definido como:

$$\xi_i = \rho \xi_{i-1} + a_i, \text{ em que:}$$

ξ_i : erro na posição i .

ξ_{i-1} : erro na posição prévia ou erro correlacionado.

ρ : coeficiente de autocorrelação.

a_i : erro independente na posição i .

A variância do processo AR(1) é dada por $\sigma_\xi^2 = \rho^2 \sigma_\xi^2 + \sigma_a^2$, a qual rearranjada fornece $\sigma_\xi^2 = \sigma_a^2 / (1 - \rho^2)$. A covariância de uma série de pontos é dada por $\sigma_\xi^2 [1 \ \rho \ \rho^2 \ \rho^3 \ \rho^4 \dots]$.

Existindo o erro extra-independente η , tem-se um novo componente de variância σ_η^2 denominado *nugget variance*. Neste caso, a covariância de uma série de pontos passa a ser dada por $[(\sigma_\xi^2 + \sigma_\eta^2) \sigma_\xi^2 \rho \ \sigma_\xi^2 \rho^2 \ \sigma_\xi^2 \rho^3 \ \sigma_\xi^2 \rho^4 \ \dots]$. Sob este novo modelo, a autocorrelação no “lag 1” passa a ser dada

por $\sigma_\xi^2 \rho / (\sigma_\xi^2 + \sigma_\eta^2)$, a autocorrelação no “lag 2” passa a ser dada por $\sigma_\xi^2 \rho^2 / (\sigma_\xi^2 + \sigma_\eta^2)$ etc. Com a inclusão de η no modelo a ser ajustado, σ_η^2 é estimado à parte e a estimativa da autocorrelação aumenta em magnitude, em relação à autocorrelação estimada com base no modelo sem o ajuste de η . Isto porque, quando σ_η^2 é omitida do modelo, a variância correlacionada tem que acomodar toda a variância. E sendo a covariância constante através dos dois modelos, tem-se as seguintes correlações no “lag 1” para os dois modelos:

(i) Modelo sem o ajuste de η

$$\rho_s = \frac{\sigma_\xi^2 \rho}{\sigma_\xi^2 + \sigma_\eta^2}$$

(ii) Modelo com o ajuste de η

$$\rho_c = \frac{\sigma_\xi^2 \rho}{\sigma_\xi^2} = \rho$$

Verifica-se que ρ_c é maior que ρ_s e que pode ser necessário incluir η no modelo.

14.19.4. Modelos lineares mistos espaciais ao nível de indivíduos com estrutura auto-regressiva de erros

A seguir é apresentado um modelo geral de estimação REML e predição BLUP com incorporação de um componente espacial.

Um modelo linear misto geral é da forma (Henderson, 1984): $y = Xb + Za + e$ (1), com as seguintes distribuições e estruturas de médias e variâncias:

$$a \sim N(0, G)$$

$$E(y) = Xb$$

$$e \sim N(0, R)$$

$$Var(y) = V = ZGZ + R$$

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.

0 : 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 .

$\hat{a} = GZV^{-1}(y - X\hat{b}) = CV^{-1}(y - X\hat{b})$: 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 elas 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:

$$\begin{aligned} L &= -\frac{1}{2} (\log|XV^{-1}X| + \log|V| + v \log \sigma_e^2 + y'Py/\sigma_e^2) \\ &= -\frac{1}{2} (\log|C^*| + \log|R| + \log|G| + v \log \sigma_e^2 + y'Py/\sigma_e^2) \end{aligned}$$

em que:

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

$v = N - r(x)$: 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.

Sendo geral, o modelo (1) contempla vários modelos inerentes às diferentes situações, tais quais o modelo geoestatístico ou de séries temporais para análise espacial, em que:

$R = \Sigma$: matriz não diagonal que considera a correlação entre resíduos, por exemplo, linhas auto-regressivas e colunas auto-regressivas ou estrutura de covariância baseada em semivariâncias ajustadas, para contemplar a autocorrelação espacial entre as observações.

Os modelos espaciais (geoestatísticos ou de séries temporais) permitem estudar a variabilidade espacial do solo nas áreas experimentais, por meio do uso de procedimentos que permitem melhor critério de estratificação ambiental (para seleção massal ou para melhor definição dos efeitos fixos no procedimen-

to BLUP). A análise espacial pode também ser realizada simultaneamente à predição BLUP dos valores genéticos e estimação REML dos componentes simplesmente incorporando $R = \Sigma$ nos estimadores e preditores.

Assim, as diferenças entre dois modelos, um espacial e outro não, são dadas apenas pela suposição associada ao erro experimental. Neste contexto, é relevante relatar que entender e modelar o erro não é tarefa fácil em Estatística e Genética Quantitativa. Entretanto, é de suma relevância uma vez que um modelo de erros escolhido aleatoriamente pode não representar a realidade e comprometer todas as inferências.

Dentre as várias formas de modelagem de Σ , o modelo de séries temporais auto-regressivo separável de primeira ordem em duas dimensões (AR1 x AR1) tem-se mostrado superior (Grondona et al., 1996; Gilmour et al., 1997; Cullis et al., 1998; Apiolaza et al., 2000; Qiao et al., 2000; Costa & Silva, 2001). Atualmente é um dos procedimentos adotados no software ASREML (Gilmour et al., 2000) desenvolvido com o objetivo de permitir análises espaciais de modelos mistos via REML/BLUP e que tem sido usado amplamente (Gilmour et al., 1997; Cullis et al., 1998; Apiolaza et al., 2000; Qiao et al., 2000; Costa & Silva, 2001) para esta finalidade. Assim, tal modelo será detalhado a seguir.

De maneira genérica, o erro é modelado em função de um efeito de tendência (ξ) mais um resíduo aleatório não correlacionado (η). Assim, o vetor de erros é particionado em $e = \xi + \eta$, em que ξ e η referem-se aos vetores de erros espacialmente correlacionados e resíduos aleatórios independentes, respectivamente. Os modelos de análise tradicionais não englobam o componente ξ e alguns modelos de análise espacial empregados em agricultura incluem ξ mas ignoram η . O resíduo η contempla efeitos genéticos não aditivos, efeitos de microambiente e erros de medida. A não modelagem de η pode ter como consequência a superestimação da variância genética aditiva. Dessa forma, tal efeito é de capital importância no ajuste de modelos mistos individuais.

Considerando um experimento com forma retangular em uma grade de c colunas e r linhas, os resíduos podem ser arranjados em uma matriz de forma que estes podem ser considerados como correlacionados dentro de linhas e colunas. Escrevendo esses resíduos em um vetor na ordem de campo pela colocação de uma coluna sob outra, a variância dos resíduos é dada por $\text{Var}(e) = \text{Var}(\xi + \eta) = R = \Sigma = \sigma_{\xi}^2 [\sum_c (\Phi_c) \otimes \sum_r (\Phi_r)] + I\sigma_{\eta}^2$, em que σ_{ξ}^2 é a variância devida a tendência é σ_{η}^2 a variância dos resíduos não correlacionados. As matrizes $\sum_c (\Phi_c)$ e $\sum_r (\Phi_r)$ referem-se a matrizes de correlação auto-regressivas de primeira ordem com parâmetros de autocorrelação Φ_c e Φ_r e ordem igual ao número de colunas e número de linhas, respectivamente. Assim, ξ é modelado como um processo auto-regressivo separável de primeira ordem (AR1 x AR1).

com matriz de covariância $Var(\xi) = \sigma_\xi^2 [\sum_c (\Phi_c) \otimes \sum_r (\Phi_r)]$ (Gilmour et al., 1997; 2000). Coeficientes de autocorrelação altos ($> 0,60$) podem indicar padrões de resíduos espaciais ou variabilidade espacial em gradientes ou em manchas. Já os coeficientes de autocorrelação baixos indicam variabilidade espacial aleatória ou sem padrão definido.

Este modelo preserva a estrutura e a informação do delineamento experimental (considere os efeitos de bloco) e pode diferir dos modelos usados em agricultura apenas por considerar σ_η^2 e por ser ao nível de indivíduos e não de parcela. Qiao et al. (2000) relataram a importância de preservar a informação do delineamento experimental e concluíram que ambos, tanto a informação do delineamento quanto o ajustamento espacial da tendência, necessitam ser considerados. Tais autores comentam que bons delineamentos experimentais e apropriados (avançados) procedimentos analíticos são importantes na obtenção de eficiência na seleção.

Tal modelo difere daquele de análise em blocos casualizados por três parâmetros a mais: σ_ξ^2 , Φ_c e Φ_r , os quais são estimados por REML. Assim, o modelo em blocos é hierárquico em relação ao espacial e a significância ou melhoramento do modelo devido à adoção da abordagem espacial pode ser avaliada pelo teste da razão de verossimilhança (LRT), confrontando duas vezes a mudança no log da função de verossimilhança maximizada contra o valor da distribuição qui-quadrado com 3 graus de liberdade, em teste bilateral. As significâncias dos vários efeitos aleatórios ajustados pelo modelo espacial podem também ser avaliadas pelo LRT. Outro critério de seleção do modelo é o “Critério de Informação de Akaike” ou AIC, o qual penaliza a verossimilhança pelo número de parâmetros ajustados independentemente. Segundo este critério, qualquer parâmetro extra (a mais) deve aumentar a verossimilhança em pelo menos uma unidade a fim de que seja incluído no modelo. O AIC é dado por $AIC = -2 \log L + 2 p$, em que L equivale a menos duas vezes o valor do log da função de verossimilhança maximizada e p é o número de parâmetros estimados. Menores valores de AIC refletem melhor ajustamento global (Akaike, 1974).

Para ajuste de modelos espaciais no ASREML, deve-se declarar a variância associada à estrutura de correlação por meio dos comandos AR, MA ou ARMA, referindo-se aos modelos auto-regressivos, médias móveis ou auto-regressivo de médias móveis, respectivamente. O comando ! units declara a inclusão do termo de erro não correlacionado.

Considerando o exemplo com 33 progênies de meios-irmãos, 6 blocos e 6 plantas por parcela, tem-se o seguinte arquivo de comandos (ignorando o efeito de parcela) para o ajuste de um modelo espacial auto-regressivo de primeira ordem em duas dimensões com um termo de erro não correlacionado:

Título
 Individuo !P
 Pai
 Mae 33
 Bloco 6
 Coluna 33
 Linha 36
 Sob
 Pedigree.txt !SKIP 1 !MAKE
 Dados.txt !SKIP 1
 Sob !BIN !LOGIT ~ Bloco !r Individuo ! units
 1 2
 Coluna Coluna AR .3
 Linha Linha AR .9

O comando 1 2 declara que o erro correlacionado deve ser modelado como um produto direto separável de 2 termos, ou seja, como $Var(\xi) = \sigma_\xi^2 [\sum_c (\Phi_c) \otimes \sum_r (\Phi_r)]$.

O comando Coluna Coluna AR .3 indica que o primeiro componente da estrutura de correlação é definido pelo fator coluna com uma matriz de correlação auto-regressiva com um valor inicial do parâmetro de correlação igual a 0,3. O segundo termo Coluna declara para o ASREML ordenar os dados, uma vez que tais podem não estar na ordem de campo.

O comando Linha Linha AR .9 indica que o segundo componente da estrutura de correlação é definido pelo fator linha com uma matriz de correlação auto-regressiva com um valor inicial do parâmetro de correlação igual a 0,9.

Em geral, a seguinte seqüência de modelos pode ser ajustada:

(a) Modelo de blocos ao acaso (Modelo 1)

Título
 Individuo 84
 Clone 14
 Bloco 2
 Parcela 28
 Peso
 Dados.txt !SKIP 1
 Peso ~ Bloco !r Clone Parcela

(b) Modelo (AR1 xAR1) sem a inclusão do termo de erro independente (Modelo 2)

Título
 Individuo 84

Clone 14
Bloco 2
Parcela 28
Coluna 14
Linha 6
Peso
Dados.txt !SKIP 1
Peso ~ Bloco !r Clone Parcela
1 2
Coluna Coluna AR .3
Linha Linha AR .9

(c) Modelo (AR1 xAR1) com a inclusão do termo de erro independente (Modelo 3)

Quando ambas as correlações auto-regressivas são significativas, o termo de erro independente pode ser acrescentado visando verificar se ocorrem melhorias no ajuste do modelo. O programa é, então, dado por:

Título
Individuo 84
Clone 14
Bloco 2
Parcela 28
Coluna 14
Linha 6
Peso
Dados.txt !SKIP 1
Peso ~ Bloco !r Clone Parcela units
1 2
Coluna Coluna AR .3
Linha Linha AR .9

(d) Modelo (AR1 xAR1) com a inclusão de linhas e colunas como efeitos aleatórios (Modelo 4)

Título
Individuo 84
Clone 14
Bloco 2
Parcela 28

Coluna 14

Linha 6

Peso

Dados.txt !SKIP 1

Peso ~ Bloco !r Clone Parcela Coluna Linha

1 2

Coluna Coluna AR .3

Linha Linha AR .9

Algumas vezes os efeitos de colunas e linhas mostram-se significativos devido a práticas diferenciadas de cultura e de colheita.

14.19.5. Aplicação a dados experimentais

Foram considerados dados referentes a um experimento de avaliação de populações, progênies e indivíduos de erva-mate para dois caracteres: massa foliar em poda de formação (peso1) e massa foliar em poda de produção (peso2). O experimento contemplou 8 populações e 141 progênies avaliadas em 10 blocos casualizados com 6 plantas por parcela, perfazendo um total de 8.460 indivíduos. Para cada caráter foram considerados dois conjuntos de dados, um contemplando os 10 blocos e outro usando apenas os 3 primeiros blocos.

Os resultados referentes ao conjunto completo de dados são apresentados nas Tabelas 14.25 (considerando o efeito de blocos como fixo) e 14.26 (considerando o efeito de blocos como aleatório). Os resultados referentes ao conjunto de 3 blocos são apresentados na Tabelas 14.27 (peso 2) e 14.28 (peso 1).

Tabela 14.25. Estimativas dos parâmetros variância genética entre progênies ($\hat{\sigma}_g^2$), entre procedências ($\hat{\sigma}_{proc}^2$), entre parcelas ($\hat{\sigma}_c^2$), residual ($\hat{\sigma}_e^2$), variância fenotípica ($\hat{\sigma}_{yi}^2$), herdabilidade individual no sentido restrito (\hat{h}_{ai}^2), herdabilidade individual entre procedências (\hat{h}_{proc}^2), coeficiente de determinação dos efeitos de parcela (\hat{c}^2), coeficientes de autocorrelação residual nas colunas (AR Coluna) e linhas (AR Linha). Dados do experimento completo.

Parâmetros	Peso 2		Peso 1	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2
Parâmetros				
$\hat{\sigma}_g^2$	0,08344	0,09396	0,01107	0,01261
$\hat{\sigma}_{proc}^2$	0,4167	0,4276	0,04428	0,04164

Continua...

Tabela 14.25. Continuação.

Parâmetros	Peso 2		Peso 1	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2
$\hat{\sigma}_c^2$	0,1385	0,0178	0,03654	0,003211
$\hat{\sigma}_e^2$	1,3390	1,4416	0,2214	0,2479
\hat{h}_{ai}^2	0,1688	0,1900	0,1413	0,1652
$\hat{h}_{proc.}^2$	0,2107	0,2160	0,1413	0,1364
$\hat{c}^2 = \hat{\sigma}_e^2 / \hat{\sigma}_{yi}^2$	0,07	0,00	0,12	0,01
Log L	-5926,86	-5712,26	1552,89	1931,14
$\hat{\sigma}_{yi}^2$	1,9776	1,9794	0,3133	0,3054
AR Coluna	---	0,2057	---	0,2877
AR Linha	---	0,1259	---	0,1399
Mudança em Log L	---	214,6**	---	378,25**
Eficiência da análise espacial (\hat{h}^2)		1,13	----	1,17
$\hat{\rho}_{gf} = (1/4)\hat{h}_{ai}^2 / [(1/4)\hat{h}_{ai}^2 + \hat{c}^2]$	0,376	----	0,2325	----
$\hat{c}^2 / \hat{h}_{ai}^2$	0,41	----	0,85	----
$\hat{\rho}_{gi} = \hat{h}_{ai}^2 / (\hat{h}_{ai}^2 + \hat{c}^2)$	0,71	----	0,54	----

Tabela 14.26. Estimativas dos parâmetros variância genética entre progêneres ($\hat{\sigma}_g^2$), entre procedências ($\hat{\sigma}_{proc.}^2$), entre parcelas ($\hat{\sigma}_c^2$), entre blocos ($\hat{\sigma}_b^2$), residual ($\hat{\sigma}_e^2$), variância fenotípica ($\hat{\sigma}_{yi}^2$), herdabilidade no sentido restrito (\hat{h}_{ai}^2), coeficiente de determinação dos efeitos de parcela (\hat{c}^2) e de blocos (\hat{b}^2), coeficientes de autocorrelação residual nas colunas (AR Coluna) e linhas (AR Linha). Dados do experimento completo.

Parâmetros	Peso 2		Peso 1	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2
$\hat{\sigma}_g^2$	0,08344	0,09397	0,01107	0,01261
$\hat{\sigma}_{proc.}^2$	0,4167	0,4277	0,04428	0,04164
$\hat{\sigma}_c^2$	0,1385	0,0178	0,03654	0,003211
$\hat{\sigma}_b^2$	0,03957	0,03804	0,007964	0,003183
$\hat{\sigma}_e^2$	1,3389	1,4417	0,2214	0,2479

Continua...

Tabela 14.26. Continuação.

Parâmetros	Peso 2		Peso 1	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2
h_{ai}^2	0,1654	0,1862	0,1378	0,1613
$\hat{c}^2 = \hat{\sigma}_e^2 / \hat{\sigma}_{yi}^2$	0,07	0,00	0,115	0,01
\hat{b}^2	0,02	0,02	0,025	0,025
Log L	-5918,26	-5703,58	1568,70	1947,17
$\hat{\sigma}_{yi}^2$	2,0171	2,0192	0,3213	0,3127
AR Coluna	---	0,2057	---	0,2877
AR Linha	---	0,1257	---	0,1401
Mudança em Log L	---	214,68*	---	378,47**
Eficiência da análise espacial (\hat{h}^2)	---	1,13	---	1,17
$\hat{\rho}_{gf} = (1/4)\hat{h}_{ai}^2 / [(1/4)\hat{h}_{ai}^2 + \hat{c}^2]$	0,37	-----	0,23	-----
$\hat{c}^2 / \hat{h}_{ai}^2$	0,42	-----	0,83	-----
$\hat{\rho}_{gi} = \hat{h}_{ai}^2 / (\hat{h}_{ai}^2 + \hat{c}^2)$	0,70	-----	0,545	-----

Verificou-se que o modelo espacial (modelo 2) apresentou melhor ajuste que o tradicional em blocos casualizados (modelo 1) para ambos os caracteres, conforme pode ser visto pelas mudanças em Log L, as quais foram altamente significativas pelo teste qui-quadrado. Constataram-se, também, aumentos das estimativas da variância genética entre progênies e do coeficiente de herdabilidade e consequentemente das eficiências seletivas que equivaleram a 1,13 e 1,17 para os caracteres peso 2 e peso 1, respectivamente (Tabelas 14.25 e 14.26).

A principal diferença entre as estimativas propiciadas pelos 2 modelos referem-se à variância ambiental entre parcelas e, consequentemente, ao componente c^2 . No modelo 1, as estimativas de c^2 foram da ordem de 0,07 e 0,12 ao passo que sob o modelo 2 estas estimativas caíram a praticamente 0% (Tabelas 14.25 e 14.26). Em outras palavras, a vantagem do modelo espacial advém da redistribuição da variação ambiental entre parcelas, na variância genética e na variância residual. Sob o modelo de blocos aleatórios (Tabela 14.26), verificou-se que a análise espacial não contribuiu para a redução da variação entre blocos e do coeficiente de determinação dos efeitos de bloco (b^2), mostrando que o modelo auto-regressivo não se mostrou capaz de remover a variação entre blocos. Este resultado revela a importância do delineamento experimental e da consideração da informação deste delineamento por ocasião da análise estatística.

Pode-se inferir que a análise espacial, via modelos auto-regressivos propicia, por análise, o que o experimentador deseja, ou seja, blocos homogêneos

sem variação ambiental entre parcelas dentro deles. A vantagem do uso deste procedimento analítico está diretamente relacionada à magnitude do coeficiente c^2 e, sobretudo, da magnitude deste em relação à herdabilidade individual, a qual pode ser avaliada por c^2/h^2 ou pelo próprio ρ_g que, no caso de testes de progêneres de polinização aberta, pode ser dado por $\rho_{gi} = h_{ai}^2/(h_{ai}^2 + c^2)$ ou $\rho_{gf} = (1/4)h_{ai}^2/[(1/4)h_{ai}^2 + c^2]$ aos níveis de indivíduo e de famílias, respectivamente. Com base nos resultados das Tabela 14.25; 14.26; 14.27 e 14.28 pode-se inferir que a análise espacial trará vantagens principalmente quando $c^2/h^2 \geq (1/3)$, ou equivalentemente, $\rho_{gi} \leq 0,75$ ou $\rho_{gf} \leq 0,43$. Isto equivale a c^2 de magnitudes iguais ou superiores a 0,05; 0,07 e 0,10 para herdabilidades individuais da ordem de 0,15; 0,20 e 0,30, respectivamente.

Tabela 14.27. Estimativas dos parâmetros variância genética entre progêneres ($\hat{\sigma}_g^2$), entre procedências ($\hat{\sigma}_{proc.}^2$), entre parcelas ($\hat{\sigma}_c^2$), residual ($\hat{\sigma}_e^2$), do erro independente ($\hat{\sigma}_\eta^2$), variância fenotípica ($\hat{\sigma}_{yi}^2$), herdabilidade individual no sentido restrito (\hat{h}_{ai}^2), herdabilidade individual entre procedências ($\hat{h}_{proc.}^2$), coeficiente de determinação dos efeitos de parcela (\hat{c}^2), variância entre colunas ($\hat{\sigma}_{col.}^2$), variância entre linhas ($\hat{\sigma}_{linha}^2$), coeficientes de autocorrelação residual nas colunas (AR Coluna) e linhas (AR Linha). Dados de apenas três repetições. Caráter peso 2.

Parâmetros	Peso2			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
$\hat{\sigma}_g^2$	0,022107	0,054938	0,066147	0,06022
$\hat{\sigma}_{proc.}^2$	0,4652	0,4632	0,4396	0,4625
$\hat{\sigma}_c^2$	0,2079	0,0000	0,0000	0,000
$\hat{\sigma}_e^2$	1,42601	1,5819	0,5768	1,5696
$\hat{\sigma}_\eta^2$	-----	-----	1,0485	-----
\hat{h}_{ai}^2	0,0417	0,1046	0,1242	0,1151
$\hat{h}_{proc.}^2$	0,2193	0,2206	0,2063	0,2210
$\hat{c}^2 = \hat{\sigma}_e^2 / \hat{\sigma}_{yi}^2$	0,098	0,00	0,00	0,00
Log L	-1881,54	-1803,23	-1750,05	-1801,03
$\hat{\sigma}_{yi}^2$	2,1212	2,1000	2,1310	2,0923
AR Coluna	-----	0,2186	0,7883	0,2120
AR Linha	-----	0,1698	0,5657	0,1701

Continua...

Tabela 14.27. Continuação.

Parâmetros	Peso 2			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Parâmetros				
Mudança em Log L	----	78,31**	131,49**	80,51**
Eficiência da análise espacial (\hat{h}^2)	----	2,51	2,98	2,76
$\hat{\sigma}_{col}^2$	----	----	----	0,000
$\hat{\sigma}_{linha}^2$	----	----	----	0,01641
$s = \hat{\sigma}_e^2 + \hat{\sigma}_\eta^2$	----	----	1,6253	----
$s = \hat{\sigma}_e^2 + \hat{\sigma}_c^2$	1,6339	1,5819	----	1,5696
$s/(\sigma_e^2 + \sigma_c^2) Model\zeta$	-----	0,9682	0,9947	0,9606
$\hat{\rho}_{gf} = (1/4)\hat{h}_{ai}^2 / [(1/4)\hat{h}_{ai}^2 + \hat{c}^2]$	0,096	----	----	----
$\hat{c}^2 / \hat{h}_{ai}^2$	2,35	----	----	----
$\hat{\rho}_{gi} = \hat{h}_{ai}^2 / (\hat{h}_{ai}^2 + \hat{c}^2)$	0,30	----	----	----

Tabela 14.28. Estimativas dos parâmetros variância genética entre progêneres ($\hat{\sigma}_g^2$), entre procedências ($\hat{\sigma}_{proc.}^2$), entre parcelas ($\hat{\sigma}_c^2$), residual ($\hat{\sigma}_e^2$), do erro independente ($\hat{\sigma}_\eta^2$), variância fenotípica ($\hat{\sigma}_{yi}^2$), herdabilidade individual no sentido restrito (\hat{h}_{ai}^2), herdabilidade individual entre procedências ($\hat{h}_{proc.}^2$), coeficiente de determinação dos efeitos de parcela (\hat{c}^2), variância entre colunas ($\hat{\sigma}_{col}^2$), variância entre linhas ($\hat{\sigma}_{linha}^2$), coeficientes de autocorrelação residual nas colunas (AR Coluna) e linhas (AR Linha). Dados de apenas três repetições. Caráter peso 1.

Parâmetros	Peso 1			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Parâmetros				
$\hat{\sigma}_g^2$	0,008516	0,01147	0,01002	0,01139
$\hat{\sigma}_{proc.}^2$	0,04439	0,04062	0,03950	0,04089
$\hat{\sigma}_c^2$	0,0326	0,0000	0,0000	0,0000
$\hat{\sigma}_e^2$	0,23865	0,2623	0,1181	0,2581
$\hat{\sigma}_\eta^2$	-----	-----	0,1507	-----
\hat{h}_{ai}^2	0,1051	0,1459	0,1259	0,1468

Continua...

Tabela 14.28. Continuação.

Parâmetros	Peso 1			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
h_{proc}^2	0,1369	0,1292	0,1241	0,1317
\hat{c}^2	0,1006	0,00	0,00	0,00
Log L	381,369	512,533	564,110	516,421
$\hat{\sigma}_{yi}^2$	0,3242	0,3144	0,3183	0,3104
AR Coluna	-----	0,2874	0,70303	0,2768
AR Linha	-----	0,1632	0,5059	0,1594
Mudança em Log L	-----	131,16**	182,74**	135,05**
Eficiência da análise espacial (\hat{h}^2)	-----	1,39	1,20	1,40
$\hat{\sigma}_{col}^2$	-----	-----	-----	0,0000
$\hat{\sigma}_{linha}^2$	----	-----	-----	0,0048
$s = \hat{\sigma}_e^2 + \hat{\sigma}_\eta^2$	-----	0,2688	-----
$s = \hat{\sigma}_e^2 + \hat{\sigma}_c^2$	0,2713	0,2623	-----	0,2581
$s/(\sigma_e^2 + \sigma_c^2) Model_f$	1,00	0,9668	0,9907	0,9513
$\hat{\rho}_{gf} = (1/4)\hat{h}_{ai}^2 / [(1/4)\hat{h}_{ai}^2 + \hat{c}^2]$	0,207	-----	-----	-----
$\hat{c}^2 / \hat{h}_{ai}^2$	0,957	-----	-----	-----
$\hat{\rho}_{gi} = \hat{h}_{ai}^2 / (\hat{h}_{ai}^2 + \hat{c}^2)$	0,51	-----	-----	-----

Os resultados das Tabelas 14.27 e 14.28 revelam que os modelos 2, 3 e 4 apresentaram ajuste significativamente melhor que o modelo 1, conforme pode ser visto pela mudança em Log L. Já as diferenças entre os ajustes dos modelos 2 e 4 tenderam a ser não significativas, mostrando que os efeitos de linhas e colunas apresentaram pequena magnitude conforme pode ser observado por $\hat{\sigma}_{col}^2$ e $\hat{\sigma}_{linha}^2$ nas Tabelas 14.27 e 14.28. As diferenças de ajuste entre os modelos 2 e 3 foram altamente significativas, revelando a necessidade de ajuste do termo de erro independente η .

Todos os três modelos de análise espacial absorveram a variação ambiental entre parcelas, tornando os valores de \hat{c}^2 iguais a zero. Verifica-se, também, que a soma das variâncias residuais ($\hat{\sigma}_e^2 + \hat{\sigma}_\eta^2$) ou ($\hat{\sigma}_e^2 + \hat{\sigma}_c^2$) foram sempre menores pelos modelos 2, 3 e 4 em relação ao modelo 1, mostrando que a abordagem espacial transformou parte da variação residual pelo modelo 1 em variação genética pelos modelos 2, 3 e 4 (Tabelas 14.27 e 14.28).

Em geral, o modelo de melhor ajuste foi o 3, o qual elevou a herdabilidade estimada de 0,04 para 0,12 e de 0,105 para 0,126 para os caracteres peso 2 e

peso 1, respectivamente, com eficiências seletivas de 2,98 e 1,20 para peso 2 e peso 1, respectivamente (Tabelas 14.27 e 14.28). É importante relatar que o modelo de melhor ajuste não é, necessariamente, aquele que fornece maior estimativa de herdabilidade, mas aquele que fornece, significativamente, maior Log L.

As estimativas dos coeficientes de autocorrelação serial foram todas positivas (Tabelas 14.25, 14.26, 14.27 e 14.28) indicando uma correlação positiva entre o crescimento de plantas vizinhas devido à similaridade ambiental. Estimativas positivas eram mesmo esperadas em erva-mate, tendo em vista que o sistema de utilização da cultura (poda bianual) minimiza os efeitos de competição, os quais tendem a gerar correlações negativas entre o crescimento de plantas vizinhas. Em espécies florestais, a competição tende a ocorrer em idades avançadas, fato que pode impactar (diminuir a magnitude) as estimativas de autocorrelação e reduzir a eficiência da análise espacial quando não se contemplar a competição nos modelos. As estimativas de autocorrelação aumentaram do modelo 2 para o modelo 3, uma vez que este último modelo ajusta o termo de erro independente e, consequentemente, diminui a variância correlacionada.

Outro fator interessante observado refere-se à obtenção das estimativas de herdabilidade quando foram usados os dados completos e apenas 3 repetições. Pelo modelo em blocos casualizados, as estimativas de herdabilidade com o dado completo equivaleram a 0,17 e 0,14 para peso 2 e peso 1, respectivamente, e com apenas 3 repetições equivaleram a 0,04 e 0,105 para peso 2 e peso 1, respectivamente. A análise espacial dos dados de 3 repetições pelo modelo 3 conduziram a estimativas de herdabilidade equivalentes a 0,124 e 0,126, portanto, bem mais próximas às estimadas com os dados completos. Isto mostra que a análise espacial pode melhorar as estimativas e minimizar um pouco os efeitos adversos da amostragem restrita, em algumas situações.

Costa e Silva et al. (2001) realizaram análises espaciais de 12 experimentos com espécies florestais (*Pinus* e *Picea*) empregando o software ASREML. Concluíram que ocorreram melhorias no modelo com a inclusão dos parâmetros espaciais e que as estimativas dos parâmetros de autocorrelação apresentaram alta magnitude (as estimativas de Φ_c e Φ_r foram superiores a 0,60). Também, Kusnadar & Galwey (2000) constataram que um modelo espacial propiciou melhor ajuste que o delineamento de blocos incompletos para dados de crescimento individual em *Pinus pinaster*.

A abordagem espacial tende a ser vantajosa, mas a sua eficiência não pode ser generalizada. Esta eficiência depende de cada situação experimental e do caráter avaliado. Recomenda-se realizar, previamente, uma análise ambiental especulativa (tópico 19.11) antes de enveredar pelas aplicações de modelos complexos. Mas tal abordagem é relevante e deve ser difundida, pois, na pior das hipóteses, produz resultados idênticos aos da análise tradicional.

Previamente à realização de análises espaciais, deve-se procurar verificar a existência ou não de uma estrutura de dependência espacial nos experimentos, ou seja, deve-se diagnosticar a presença de variabilidade espacial. E isto pode ser verificado de maneira simples com base nos parâmetros c^2 e ρ_g . Valores altos de c^2 indicam alta variabilidade entre parcelas dentro de blocos e alta correlação ambiental entre observações dentro de parcela. Por sua vez, valores baixos de ρ_g indicam que os tratamentos são ordenados diferentemente de um bloco para outro, fato que pode ser devido a um gradiente de fertilidade, segundo Eisenberg et al. (1996).

Outros testes estatísticos simples destinados à avaliação da autocorrelação ou correlação serial de resíduos em análise de séries temporais ou de regressão, como o teste de Durbin-Watson, podem também ser aplicados na avaliação da dependência espacial. O teste de Durbin-Watson é dado por (Morettin & Toloi, 1987):

$$dw = \frac{\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^n (\hat{e}_i^2)}, \text{ em que } i = 1, 2, \dots, n \text{ refere-se à ordem da posição da}$$

observação vinculada ao erro \hat{e}_i . No caso, \hat{e}_i e \hat{e}_{i-1} são erros de observações em que as posições são adjacentes, ou seja, de observações com vizinhança de primeira ordem.

A estatística dw apresenta relação direta com a autocorrelação de primeira ordem (ρ), dada por $dw = 2(1 - \rho)$. Assim, $\rho = (2-dw)/2$ e, em ausência de autocorrelação espacial, dw tem valor esperado igual a 2. Valores de dw superiores a 2 indicam autocorrelação negativa ao passo que valores de dw inferiores a 2, autocorrelação positiva. Em geral, a heterogeneidade espacial causa autocorrelações positivas, ao passo que os efeitos de competição entre plantas vizinhas causam autocorrelações negativas. Maiores detalhes sobre análise espacial de experimentos são apresentados por Duarte (2000) e Resende & Sturion (2001), incluindo também métodos geoestatísticos.

Capítulo 15

**Delineamentos de Cruzamento
e Comparação de Métodos
de Melhoramento**

15.1. Introdução

Os delineamentos de cruzamento são extremamente importantes no contexto das estratégias de melhoramento genético, pois estão relacionados com a eficiência seletiva nos testes de progênies. Os principais objetivos dos testes de progênies são (Bridgewater, 1992): a) avaliação do valor genético dos genitores; b) estimativa de parâmetros genéticos; c) predição de ganhos genéticos realizados; d) geração da população-base para nova seleção. Assim, os delineamentos de cruzamento devem ser estabelecidos com base nesses objetivos.

Burdon (1992) resume os propósitos dos delineamentos de cruzamento em duas categorias principais: avaliação genética de genitores e geração de população-base para nova seleção. Esta denotação é adequada uma vez que, para atender de forma adequada aos objetivos dos itens a) e d) acima, necessariamente, atende-se aos itens b) e c). Para avaliação da capacidade geral de combinação dos genitores, um número de 4 a 5 cruzamentos com outros genitores tem sido considerado adequado (Burdon, 1992). Os seguintes requisitos devem ser observados na geração de uma população-base adequada: (i) grande número de indivíduos e razoável número de cruzamentos para propiciar alta intensidade de seleção; (ii) possibilidade de gerenciamento do parentesco entre indivíduos por meio de informações de genealogia, geradas por cruzamentos controlados; (iii) tamanho efetivo alto; (iv) garantia de que cada genitor não tenha sido cruzado apenas com genitores bons ou ruins. Este último requisito é importante, sobretudo, para evitar que os melhores genitores sejam cruzados apenas com genitores ruins, pois cruzamentos deste tipo conduzirão à não seleção das progênies geradas por estes genitores melhores. Neste sentido, alguma forma de seleção de cruzamentos é desejável.

15.2. Principais Delineamentos de Cruzamento

Na Tabela 15.1. são descritas as principais características de alguns delineamentos de cruzamento. Para descrição prática dos diferentes delineamentos ver o trabalho de Shimizu et al. (1982) e Bridgewater (1992).

Tabela 15.1. Delineamentos de cruzamento e suas principais características.

Delineamento	Nº de cruzamentos	Nº de cruzamentos para n=60 e m=4	Controle genealógico	Avaliação dos genitores	Estimação de efeitos aditivos	Estimação de efeitos não aditivos	Eliminação de cruzamentos entre genitores melhores e piores	N _e
1. Polinização aberta	n	60	N	S	S	N	N	A
2. Policruzamento (meios-irmãos)	n	60	N	S	S	N	S	A

Continua...

Tabela 15.1. Continuação.

Delineamento	Nº de cruzamentos	Nº de cruzamentos para n=60 e m=4	Controle genealógico	Avaliação dos genitores	Estimação de efeitos aditivos	Estimação de efeitos não aditivos	Eliminação de cruzamentos entre genitores melhores e piores	N _e
3. Pares simples (irmãos germanos)	n/2	30	S	N	N	N	N	A
	n/(m+1) e							
4. Hierárquico	$\frac{n \cdot m}{m + 1}$	12 e 48	S	P	P	S	S	A
5. Fatorial	n e $(n/2)^2$	60 e 900	S	S	S	S	S	A
6. Testador	(n-m) e (n-m) m	56 e 224	S	S	S	P	S	B
7. Dialélico	n e n^2	60 e 3.600*	S	S	S	S	S	A
8. Meio dialélico	n e [n(n-1)]/2	60 e 1.770	S	S	S	S	S	A
9. Dialélico parcial	n e 4n	60 e 240	S	S	S	S	S	A
10. Fatorial desconexo	n e nm/2	60 e 120	S	S	S	S	S	A
11. Meio dialélico desconexo	n e n(m-1)/2	60 e 150*	S	S	S	S	S	A
12. Autofecundação	n*	60	S	S	P	N	S	B
13. Cruzamentos preferenciais positivos	V	V	S	S	S	S	S	B
14. Complementares (Testador ou policruzamento com meio dialélico desconexo ou fatorial desconexo; policruzamento e pares simples)	V	V	S	S	S	S		A

n - número de genitores na população de melhoramento; m - número de genitores no grupo menos numeroso; n/(m+1) e nm/(m+1) - número de progênies de meios-irmãos e irmãos germanos respectivamente; n* - número de progênies autofecundadas; 3.600* = 3.540 progênies de irmãos germanos e 60 progênies de autofecundação; 150* - considerando m = 6 genitores; S, N, P, A, B e V: sim, não, parcial, alto, baixo e variável, respectivamente.

Os diversos delineamentos apresentam diferentes custos operacionais e eficiências. Os mais restritivos tecnicamente são os delineamentos de polinização aberta, policruzamento, pares simples e hierárquico, principalmente para gerações avançadas de melhoramento. Os delineamentos dialélico, fatorial e meio dialélico são restritivos quanto ao número total de cruzamentos. Os delineamentos de testadores, cruzamentos preferenciais positivos e autofecundação geram populações-base muito aparentadas, com baixo tamanho efetivo popula-

cional. Dessa forma, os delineamentos mais recomendados são o dialélico parcial, o fatorial desconexo e o meio dialélico desconexo.

A otimização de diferentes delineamentos de cruzamento para seleção de genitores e seleção de indivíduos em testes de progênies pode ser realizada por meio do uso de dois ou mais delineamentos complementares, conforme abordagem proposta por Burdon & Shelbourne (1971) e Van Buijtenen (1972; 1976). Atualmente, estes delineamentos complementares têm sido bastante usados, com ênfase em um tipo de delineamento para a seleção de genitores e em outro para a criação da população-base para seleção (Burdon & Van Buijtenen, 1990; Van Buijtenen & Burdon, 1990).

A necessidade desses delineamentos complementares decorre da inevitável endogamia que surge nas populações de melhoramento (tipicamente sublinhas) em geração mais avançada, fato que mascara as inferências sobre os valores genéticos dos genitores (por meio de variação nos níveis de endogamia e/ou reposta dos genitores à endogamia). Para a avaliação de genitores, Resende & Vencovsky (1992; 1993) demonstraram que a utilização de progênies de autofecundação é muito vantajosa. Entretanto, este delineamento conduz a um rápido aumento do parentesco nas populações-base de melhoramento.

Para estimação de parâmetros genéticos, Huber et al. (1992) concluíram que o delineamento com progênies de meios-irmãos (policruzamento ou polinização aberta) conduz a estimativas mais precisas da herdabilidade no sentido restrito.

A eficiência dos programas de melhoramento florestal depende da utilização de acurados métodos de seleção e de eficientes delineamentos de cruzamento. Dentre os trabalhos relevantes sobre a comparação de delineamentos de cruzamento, destacam-se o de Cotterill (1986) e o de Van Buijtenen & Burdon (1990), os quais consideraram apenas os valores genéticos aditivos. Em termos gerais, estes trabalhos concluíram pela superioridade dos delineamentos de meio dialélico desconexo (Cotterill, 1986; Van Buijtenen, 1990) e fatorial desconexo (Van Buijtenen & Burdon, 1990), para a seleção de indivíduos em testes de progênies visando ao avanço de gerações. Van Buijtenen & Burdon (1990) constataram também a superioridade do delineamento fatorial sobre o dialélico, na situação em que cada genitor participa de um pequeno número de cruzamentos.

15.3. Comparação entre Delineamentos de Cruzamento

Na seqüência serão comparados (com base em Resende, 1999b) alguns delineamentos de cruzamento fatorial e dialélico com os de pares simples (progênies de irmãos germanos) e policruzamento (progênies de meios-irmãos) com vistas a seleção de indivíduos visando a propagação sexuada (valores genéticos aditivos) e assexuada (valores genotípicos), em programas de seleção intrapopulacional.

15.3.1. Modelos estatísticos

O modelo estatístico associado à avaliação de p progênies de meios-irmãos ou de irmãos germanos no delineamento de blocos ao acaso equivale a:

$$Y_{ijk} = \mu + p_i + b_j + e_{ij} + \delta_{ijk}, \text{ em que:}$$

$\mu = \bar{Y} \dots$: efeito da média geral, fixo.

$p_i = (\bar{Y}_i \dots - \bar{Y} \dots)$: efeito de progênies, aleatório.

$b_j = (\bar{Y}_{.j} \dots - \bar{Y} \dots)$: efeito de blocos, fixo.

$e_{ij} = (\bar{Y}_{ij} \dots - \bar{Y}_i \dots - \bar{Y}_{.j} \dots + \bar{Y} \dots)$: efeito de parcela, aleatório.

$\delta_{ijk} = (Y_{ijk} - \bar{Y}_{ij} \dots)$: efeito do indivíduo dentro de parcela, aleatório.

O modelo estatístico associado à avaliação (no delineamento de blocos ao acaso) de progênies de irmãos germanos obtidas sob cruzamentos fatoriais desconexos equivale a:

$$Y_{ijk\ell n} = \mu + s_i + b_j + f_{ik} + m_{i\ell} + fm_{ik\ell} + sb_{ij} + e_{ijkl} + \delta_{ijk\ell n}, \text{ em que:}$$

$\mu = \bar{Y} \dots \dots$: efeito da média geral, fixo.

$s_i = (\bar{Y}_i \dots \dots - \bar{Y} \dots \dots)$: efeito de grupo de genitores, aleatório.

$b_j = (\bar{Y}_{.j} \dots \dots - \bar{Y} \dots \dots)$: efeito de blocos, fixo.

$f_{ik} = (\bar{Y}_{i \cdot k} \dots \dots - \bar{Y} \dots \dots)$: efeito da capacidade geral de combinação do genitor k dentro do fatorial i , aleatório (efeito de fêmea dentro de grupo).

$m_{i\ell} = (\bar{Y}_{i \cdot \ell} \dots \dots - \bar{Y} \dots \dots)$: efeito da capacidade geral de combinação do genitor ℓ dentro do fatorial i , aleatório (efeito de macho dentro de grupo)

$fm_{ik\ell} = (\bar{Y}_{i \cdot k \cdot \ell} \dots \dots - \bar{Y}_{i \cdot k} \dots \dots - \bar{Y}_{i \cdot \ell} \dots \dots + \bar{Y} \dots \dots)$: efeito da capacidade específica de combinação entre os genitores k e ℓ dentro do fatorial i , aleatório.

$sb_{ij} = (\bar{Y}_{ij} \dots \dots - \bar{Y} \dots \dots)$: efeito da interação grupos x blocos, aleatório.

$e_{ijkl} = (\bar{Y}_{ijkl} \dots \dots - \bar{Y}_{i \cdot k \cdot \ell} \dots \dots - \bar{Y}_{.j} \dots \dots + \bar{Y} \dots \dots)$: efeito da parcela $ijkl$, aleatório.

$\delta_{ijk\ell n} = (Y_{ijk\ell n} - \bar{Y}_{ijkl} \dots \dots)$: efeito do indivíduo dentro de parcela, aleatório.

No delineamento de cruzamentos fatoriais dentro de um grupo de genitores, as covariâncias entre os efeitos do modelo estatístico e os valores genéticos aditivos e de dominância foram apresentadas no Capítulo 7, tópico 7.7. No caso de cruzamentos fatoriais com vários grupos de genitores, as demais covariâncias são relatadas a seguir.

Covariância entre efeitos aditivos (a) e efeitos fenotípicos:

$$\begin{aligned} COV[a, (\bar{Y}_i \dots - \bar{Y} \dots)] &= \frac{t-1}{t} \frac{1}{nbmf} [(1/2) \sigma_a^2 + nbm (1/4) \sigma_a^2 + nbf (1/4) \sigma_a^2 + \\ &\quad + nb (1/2) \sigma_a^2 - nb (1/2) \sigma_a^2] \\ &= \frac{t-1}{t} \left[\frac{(1/4) \sigma_a^2}{f} + \frac{(1/4) \sigma_a^2}{m} + \frac{(1/2) \sigma_a^2}{nbmf} \right] \\ &= \frac{t-1}{t} \frac{(m+f) nb+2}{4nbmf} \sigma_a^2 \\ COV[A, (\bar{Y}_i \dots - \bar{Y} \dots)] &= \frac{t-1}{t} \frac{b-1}{b} \frac{(1/2)}{nfm} \sigma_a^2 \end{aligned}$$

Covariância entre efeitos de dominância (d) e efeitos fenotípicos:

$$\begin{aligned} COV[d, (\bar{Y}_i \dots - \bar{Y} \dots)] &= \frac{t-1}{t} \frac{1}{nbmf} [\sigma_d^2 + nb (1/4) \sigma_d^2 - (1/4) \sigma_d^2] \\ &= \frac{t-1}{t} \left[\frac{(1/4) \sigma_d^2}{mf} + \frac{(3/4) \sigma_d^2}{nbmf} \right] \\ &= \frac{t-1}{t} \frac{nb+3}{4nbmf} \sigma_d^2 \\ COV[d, (\bar{Y}_i \dots - \bar{Y} \dots)] &= \frac{t-1}{t} \frac{b-1}{b} \frac{(3/4)}{nfm} \sigma_d^2, \end{aligned}$$

em que t , n , b , f e m referem-se aos números de subconjuntos, plantas por parcela, blocos, genitores femininos e genitores masculinos, respectivamente.

Com base nestes estimadores, verifica-se que o efeito $sb_{ij} = (\bar{Y}_{ij} \dots - \bar{Y} \dots)$ é negligenciável, pois apresenta covariância praticamente nula com os efeitos aditivos e de dominância. Já o efeito de subconjuntos deve ser considerado na predição dos efeitos aditivos e de dominância, especialmente para baixos valores de m e f .

A variância fenotípica associada ao efeito de subconjuntos é dada por:

$$V(6) = \frac{t-1}{t} \left[\frac{\sigma_\delta^2}{mfnb} + \frac{\sigma_e^2}{mfb} + \frac{\sigma_{fm}^2}{fm} + \frac{\sigma_f^2}{f} + \frac{\sigma_m^2}{m} + \frac{\sigma_{sb}^2}{b} + \sigma_s^2 \right].$$

Considerando a alocação aleatória de genitores nos subconjuntos, pode-se considerar $\sigma_s^2 = 0$ (Van Buijtenen & Burdon, 1990). Se, $\sigma_{sb}^2/b = 0$, tem-se:

$V(6) = \frac{t-1}{t} \left[\frac{\sigma_\delta^2}{mfnb} + \frac{\sigma_e^2}{mfb} + \frac{\sigma_{fm}^2}{fm} + \frac{\sigma_f^2}{f} + \frac{\sigma_m^2}{m} \right]$, que expressa, em termos de coeficientes de correlação intraclassse equivale a:

$$V(6) = \frac{t-1}{t} \left[\frac{1 + (n-1)t_1 + (b-1)nt_2 + (m+f-2)nbt_3}{mfnb} \sigma_y^2 \right], \text{ em que:}$$

$$\sigma_f^2 = (1/4) \sigma_a^2; \quad \sigma_m^2 = (1/4) \sigma_a^2; \quad \sigma_{fm}^2 = (1/4) \sigma_d^2; \quad \dots$$

$$\sigma_y^2 = \sigma_a^2 + \sigma_d^2 + \sigma_e^2 + \sigma_{ad}^2 = \sigma_f^2 + \sigma_m^2 + \sigma_{fm}^2 + \sigma_e^2 + \sigma_d^2 : \text{variância fenotípica.}$$

$t_1 = c^2 + t_2$ = correlação intraclasse entre os indivíduos na parcela.

$t_2 = \rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)$: correlação intraclasse entre irmãos em diferentes parcelas.

$t_3 = (1/4) h_a^2$: correlação intraclasse entre meios-irmãos paternos ou maternos em diferentes parcelas, obtidos sob cruzamento fatorial.

$c^2 = \frac{\sigma_e^2}{\sigma_y^2}$ = correlação intraclasse devida ao ambiente comum da parcela.

$h_a^2 = \frac{\sigma_a^2}{\sigma_y^2}$: herdabilidade individual no sentido restrito.

h_g^2 : herdabilidade individual no sentido amplo.

ρ_a e ρ_d : correlação genética aditiva e de dominância entre irmãos, respectivamente.

15.3.2. Índice multiefeitos para progênieis de meios-irmãos e irmãos germanos

O índice para ordenamento dos efeitos genéticos aditivos dos candidatos à seleção visando ao plantio via propagação sexuada equivale a (Resende & Higa, 1994a):

$$I_1 = b_1(Y_{ijk} - \bar{Y}_{ij.}) + b_2(\bar{Y}_{i..} - \bar{Y}_{...}) + b_3(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}), \text{ em que:}$$

$$b_1 = \frac{(1-\rho_a) h_a^2}{(1-t_1)}; \quad b_2 = \frac{[1+(nb-1)\rho_a] h_a^2}{1+(n-1) t_1 + (b-1) nt_2}; \quad b_3 = \frac{(1-\rho_a) h_a^2}{1+(n-1) t_1 - nt_2};$$

$$t_1 = c^2 + t_2; \quad t_2 = \rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)$$

No caso do ordenamento dos efeitos genotípicos dos candidatos à seleção para o plantio via propagação assexuada, o índice é equivalente a:

$$I_2 = (b_1 + b_4)(Y_{ijk} - \bar{Y}_{ij.}) + (b_2 + b_5)(\bar{Y}_{i..} - \bar{Y}_{...}) + (b_3 + b_6)(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}), \text{ em que:}$$

$$b_4 = \frac{(1-\rho_d) (h_g^2 - h_a^2)}{(1-t_1)}; \quad b_5 = \frac{[1+(nb-1)\rho_d] (h_g^2 - h_a^2)}{1+(n-1) t_1 + (b-1) nt_2}; \quad b_6 = \frac{(1-\rho_d) (h_g^2 - h_a^2)}{1+(n-1) t_1 - nt_2};$$

A acurácia associada à seleção pelo índice I_1 é igual a:

$$r_{I_1} = [\alpha_1 b_1 + \alpha_2 b_2 + \alpha_3 b_3]^{1/2}, \text{ em que:}$$

$$\alpha_1 = (1-\rho_a) \frac{n-1}{n}; \quad \alpha_2 = \frac{1+(nb-1) \rho_a}{nb} \frac{p-1}{p}; \quad \alpha_3 = \frac{(1-\rho_a)}{n} \frac{p-1}{p} \frac{b-1}{b}.$$

Já para o índice I_2 , a acurácia é:

$$r_{I_{g_2}} = [\alpha_1 b_1 + \alpha_2 b_2 + \alpha_3 b_3 + \alpha_4 b_4 + \alpha_5 b_5 + \alpha_6 b_6]^{1/2}, \text{ em que:}$$

$$\alpha_4 = (1 - \rho_d) \frac{n-1}{n}; \quad \alpha_5 = \frac{1+(nb-1)}{nb} \rho_d \frac{p-1}{p}; \quad \alpha_6 = \frac{(1-\rho_d)}{n} \frac{p-1}{p} \frac{b-1}{b}.$$

Nas expressões apresentadas:

$$\rho_d = 0,25 \text{ para meios-irmãos e } 0,5 \text{ para irmãos germanos.}$$

$$\rho_d = 0 \text{ para meios-irmãos e } 0,25 \text{ para irmãos germanos.}$$

15.3.3. Índice multiefeitos associados a cruzamentos fatoriais

Os índices para ordenamento dos efeitos genéticos aditivos e genotípicos dos candidatos à seleção para o plantio por propagação sexuada e assexuada, respectivamente, equivalem a:

Efeitos genéticos aditivos - propagação sexuada

$$I_3 = b_{11}(\bar{Y}_{i..k..} - \bar{Y}_{...}) + b_{12}(\bar{Y}_{i..l..} - \bar{Y}_{...}) + b_{13}(\bar{Y}_{i..k..l..} - \bar{Y}_{i..k..} - \bar{Y}_{i..l..} + \bar{Y}_{...}) + \\ + b_{14}(\bar{Y}_{ijk..} - \bar{Y}_{i..k..} - \bar{Y}_{j..} + \bar{Y}_{...}) + b_{15}(Y_{ijk..n} - \bar{Y}_{ijk..}) + b_{16}(\bar{Y}_{i...} - \bar{Y}_{...}),$$

em que:

$$b_{11} = \frac{\left(\frac{nbm+2}{4} \right) h_a^2}{1 + (n-1) t_1 + (b-1) nt_2 + (m-2) nbt_3};$$

$$b_{12} = \frac{\left(\frac{nbf+2}{4} \right) h_a^2}{1 + (n-1) t_1 + (b-1) nt_2 + (f-2) nbt_3}$$

$$b_{13} = \frac{(1/2) h_a^2}{1 + (n-1) t_1 + (b-1) nt_2 - 2nbt_3}$$

$$b_{14} = \frac{(1/2) h_a^2}{1 + (n-1) t_1 - nt_2}$$

$$b_{15} = \frac{(1/2) h_a^2}{(1-t_1)}$$

$$b_{16} = \frac{\left(\frac{(m+f) nb+2}{4} \right) h_a^2}{1 + (n-1) t_1 + (b-1) nt_2 + (m+f-2) nbt_3}$$

$$t_1 = c^2 + t_2; \quad t_2 = (1/2) h_a^2 + (1/4)(h_g^2 - h_a^2); \quad t_3 = (1/4) h_a^2$$

Efeitos genotípicos - propagação assexuada:

$$I_4 = (b_{11} + b_{17})(\bar{Y}_{i..} - \bar{Y}_{...}) + (b_{12} + b_{18})(\bar{Y}_{i..} - \bar{Y}_{...}) + (b_{13} + b_{19})(\bar{Y}_{i..} - \bar{Y}_{i..} - \bar{Y}_{i..} + \bar{Y}_{...}) + (b_{14} + b_{20})(\bar{Y}_{ijk..} - \bar{Y}_{i..} - \bar{Y}_{j..} + \bar{Y}_{...}) + (b_{15} + b_{21})(\bar{Y}_{ijk..} - \bar{Y}_{ijk..}) + (b_{16} + b_{22})(\bar{Y}_{i..} - \bar{Y}_{...}),$$

em que:

$$b_{17} = \frac{\left(\frac{nb+3}{4}\right)(h_g^2 - h_a^2)}{1 + (n-1)t_1 + (b-1)nt_2 + (m-2)nbt_3};$$

$$b_{18} = \frac{\left(\frac{nb+3}{4}\right)(h_g^2 - h_a^2)}{1 + (n-1)t_1 + (b-1)nt_2 + (f-2)nbt_3}$$

$$b_{19} = \frac{[1 + (nb-1)(1/4)](h_g^2 - h_a^2)}{1 + (n-1)t_1 + (b-1)nt_2 - 2nbt_3}$$

$$b_{20} = \frac{(3/4)(h_g^2 - h_a^2)}{1 + (n-1)t_1 - nt_2}$$

$$b_{21} = \frac{(3/4)(h_g^2 - h_a^2)}{(1-t_1)}$$

$$b_{22} = \frac{\frac{nb+3}{4}(h_g^2 - h_a^2)}{1 + (n-1)t_1 + (b-1)nt_2 + (m+f-2)nbt_3}$$

As acuráncias associadas à seleção pelos índices I_3 e I_4 equivalem, respectivamente, a:

Índice I_3 :

$$r_{I_{g_3}} = [\alpha_{11}b_{11} + \alpha_{12}b_{12} + \alpha_{13}b_{13} + \alpha_{14}b_{14} + \alpha_{15}b_{15}]^{1/2}, \text{ em que:}$$

$$\alpha_{11} = \left[\frac{nbm+2}{4nbm} \right] \frac{f-1}{f}; \quad \alpha_{12} = \left[\frac{nbf+2}{4nbf} \right] \frac{m-1}{m}; \quad \alpha_{13} = \frac{(1/2)}{nb} \frac{f-1}{f} \frac{m-1}{m};$$

$$\alpha_{14} = \frac{(1/2)}{n} \frac{fm-1}{fm} \frac{b-1}{b}; \quad \alpha_{15} = (1/2) \frac{n-1}{n}; \quad \alpha_{16} = \frac{(m+f)}{4nbmf} \frac{nb+2}{t} \frac{t-1}{t}.$$

Índice I_4

$$r_{I_{g_4}} = [\alpha_{11}b_{11} + \alpha_{12}b_{12} + \alpha_{13}b_{13} + \alpha_{14}b_{14} + \alpha_{15}b_{15} + \alpha_{16}b_{16} + \alpha_{17}b_{17} + \alpha_{18}b_{18} + \alpha_{19}b_{19} + \alpha_{20}b_{20}]^{1/2}, \text{ em que:}$$

$$\alpha_{17} = \frac{nb+3}{4nbf} \frac{f-1}{f}; \quad \alpha_{18} = \frac{nb+3}{4nbf} \frac{m-1}{m}; \quad \alpha_{19} = \frac{1+(nb-1)(1/4)}{nb} \frac{f-1}{f} \frac{m-1}{m};$$

$$\alpha_{20} = \frac{(3/4)}{n} \frac{fm-1}{fm} \frac{b-1}{b}; \quad \alpha_{21} = (3/4) \frac{n-1}{n}; \quad \alpha_{22} = \frac{nb+3}{4nbf} \frac{t-1}{t}.$$

15.3.4. Parâmetros assumidos na comparação

A eficiência dos diferentes delineamentos de cruzamento pode ser comparada empregando-se a expressão do progresso genético: $G_s = k r_{la} \sigma_a$ para sistemas de propagação sexuada e $G_s = k r_{lg} \sigma_g$ para sistemas de propagação assexuada. Entretanto, uma vez que σ_a , σ_g e k são constantes nos diferentes delineamentos, as eficiências relativas podem ser computadas empregando-se apenas os estimadores do parâmetro acurácia (r_{la} ou r_{lg}).

A acurácia depende dos parâmetros h_a^2 , h_g^2 e c^2 e das constantes experimentais n , p , b , m e f . Fixando a população experimental em 6.000 plantas, impõe-se a restrição de custo experimental fixo associado aos diferentes delineamentos. Com esta condição, estipulou-se $p = 100$ progénies, $n = 6$ plantas por parcela e $b = 10$ blocos. Considerou-se também $m = f = 4$ (6 fatoriais 4×4) e, alternativamente $m = f = 5$ (4 fatoriais 5×5), $m = f = 3$ (11 fatoriais 3×3) e $m = f = 2$ (25 fatoriais 2×2). Fatoriais com essas dimensões são comumente empregados na prática (Bridgewater, 1992). Alternativamente, considerou-se a estrutura experimental com uma planta por parcela e, neste caso, $n = 1$ e $b = 60$.

Assim, para comparação entre os três delineamentos de cruzamento foram considerados diferentes níveis dos parâmetros h_a^2 , h_g^2 e c^2 , conforme apresentado na Tabela 15.2.

Tabela 15.2. Valores dos parâmetros herdabilidade individual no sentido restrito (h_a^2), herdabilidade individual no sentido amplo (h_g^2) e correlação devida ao ambiente comum (c^2) utilizados na simulação.

h_a^2	h_g^2	c^2 para meios-irmãos	c^2 para irmãos germanos e fatorial	c^2 para delineamentos com 1 planta por parcela
0,05	0,05	0,0125	0,0175	0,00
	0,075	0,0125	0,02	0,00
0,10	0,10	0,025	0,035	0,00
	0,15	0,025	0,04	0,00
0,20	0,20	0,05	0,07	0,00
	0,30	0,05	0,08	0,00
0,30	0,30	0,075	0,105	0,00
	0,45	0,075	0,12	0,00
0,40	0,40	0,10	0,14	0,00
	0,60	0,10	0,16	0,00
0,50	0,50	0,125	0,175	0,00
	0,75	0,125	0,20	0,00

15.3.5. Resultados e inferências práticas

Os resultados comparativos referentes aos fatoriais desconexos 2 x 2, 3 x 3, 4 x 4 e 5 x 5, para vários valores de herdabilidade e sistemas de propagação são apresentados nas Tabelas 15.3 e 15.4.

Tabela 15.3. Acurárias seletivas associadas aos delineamentos de cruzamento fatoriais desconexos 2 x 2 (r_{IaF_2}), 3 x 3 (r_{IaF_3}), 4 x 4 (r_{IaF_4}) e 5 x 5 (r_{IaF_5}) avaliados no delineamento de blocos ao acaso com 6 plantas por parcela. Sistema de propagação: sexuada.

$h_a^2 *$	$h_g^2 *$	r_{IaF_2}	r_{IaF_3}	r_{IaF_4}	r_{IaF_5}
0,05	0,05	0,58	0,61	0,62	0,63
	0,075	0,55	0,58	0,59	0,61
	0,10	0,64	0,66	0,66	0,66
	0,15	0,61	0,62	0,64	0,64
	0,20	0,70	0,70	0,71	0,71
	0,30	0,65	0,67	0,68	0,68
	0,30	0,73	0,74	0,74	0,74
	0,45	0,69	0,71	0,71	0,71
	0,40	0,77	0,77	0,77	0,77
	0,60	0,73	0,75	0,75	0,75
0,50	0,50	0,81	0,81	0,81	0,81
	0,75	0,78	0,80	0,80	0,80

* h_a^2 e h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

Tabela 15.4. Acurárias seletivas associadas aos delineamentos de cruzamento fatoriais desconexos 2 x 2 (r_{IaF_2}), 3 x 3 (r_{IaF_3}), 4 x 4 (r_{IaF_4}) e 5 x 5 (r_{IaF_5}) avaliados no delineamento de blocos ao acaso com 6 plantas por parcela. Sistema de propagação: assexuada.

$h_a^2 *$	$h_g^2 *$	r_{IaF_2}	r_{IaF_3}	r_{IaF_4}	r_{IaF_5}
0,05	0,05	0,58	0,61	0,62	0,63
	0,075	0,60	0,63	0,65	0,66
	0,10	0,64	0,66	0,66	0,66
	0,15	0,67	0,70	0,71	0,72
	0,20	0,70	0,70	0,71	0,71
	0,30	0,75	0,77	0,79	0,80
	0,30	0,73	0,74	0,74	0,74
	0,45	0,81	0,84	0,85	0,86
	0,40	0,77	0,77	0,77	0,77
	0,60	0,89	0,91	0,92	0,93
0,50	0,50	0,81	0,81	0,81	0,81
	0,75	0,97	0,99	1,00	1,00

* h_a^2 e h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

Verifica-se que o aumento da eficiência seletiva com o aumento da dimensão do fatorial decresce rapidamente. Para caracteres com herdabilidade mais alta ($>0,30$) o ponto de eficiência máxima é praticamente atingido com o fatorial 3×3 , tanto para a propagação vegetativa quanto para a propagação sexuada. Para caracteres com herdabilidade mais baixa ($<0,20$), a maximização da eficiência seletiva praticamente é atingida com o fatorial 4×4 . Em algumas situações específicas, o fatorial 5×5 propicia uma ligeira vantagem sobre o fatorial 4×4 (Tabelas 15.3 e 15.4).

É importante verificar que, com o aumento da dimensão do fatorial, diminui o número de genitores a serem avaliados considerando um número fixo de famílias. Assim, fixando o número de famílias em 100, com os fatoriais 4×4 e 5×5 são utilizados 50 e 40 genitores, respectivamente. Desta forma, os fatoriais 4×4 podem ser preferidos por permitirem maior tamanho efetivo populacional na população de seleção, fato que pode permitir maior diferencial de seleção realizado (Resende, 1994). As vantagens adicionais dos fatoriais 4×4 sobre fatoriais menores (2×2 e 3×3) e sobre progênieis de irmãos germanos ou cruzamentos em pares simples (fatoriais 1×1 ou meio dialélico 1×1) são: (i) evitar que determinados genitores sejam cruzados apenas com genitores bons ou apenas com genitores ruins; (ii) permitir avaliar a capacidade geral de combinação e a prática da seleção de genitores (Burdon, 1992).

Na Tabela 15.5 são apresentados os resultados referentes à inclusão e à exclusão do efeito de grupo de genitores em fatoriais desconexos 3×3 .

Tabela 15.5. Acurárias associadas ao delineamento fatorial 3×3 desconexo, considerando a inclusão e a exclusão do efeito de grupo de genitores no índice de seleção. Delineamento experimental de blocos ao acaso com 6 plantas por parcela.

$h_a^2 *$	$h_g^2 *$	Sem efeito de grupo		Com efeito de grupo	
		Propagação sexuada	Propagação assexuada	Propagação sexuada	Propagação assexuada
$h_a^2 *$	$h_g^2 *$				
0,05	0,05	0,49	0,49	0,61	0,61
	0,075	0,47	0,53	0,58	0,63
0,10	0,10	0,55	0,55	0,66	0,66
	0,15	0,52	0,60	0,62	0,70
0,20	0,20	0,60	0,60	0,70	0,70
	0,30	0,56	0,68	0,67	0,77
0,30	0,30	0,63	0,63	0,74	0,74
	0,45	0,60	0,75	0,71	0,84
0,40	0,40	0,67	0,67	0,77	0,77
	0,60	0,65	0,83	0,75	0,91
0,50	0,50	0,71	0,71	0,81	0,81
	0,75	0,71	0,92	0,80	0,99

* h_a^2 e h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

Constata-se que a inclusão do efeito de grupo de genitores é fundamental em fatoriais desconexos de pequena dimensão (Tabela 15.5). A exclusão do efeito de grupo de genitores somente se justificaria quando ocorre a adoção de intensidades de seleção fixas dentro de cada fatorial.

Dentre os delineamentos de cruzamento mais elaborados, três são mais indicados e utilizados no melhoramento florestal por permitirem a utilização de um número adequado de genitores por meio da geração de um número não muito elevado de famílias de irmãos germanos: o fatorial desconexo, o dialélico parcial e o meio dialélico desconexo.

Os dialélicos parciais são muito similares aos fatoriais e alguns podem ser convertidos em fatoriais. Por exemplo, o dialélico parcial

1	1	2	3	4	5	6	7	8
1		X		X		X		X
2			X		X		X	
3				X		X		X
4					X		X	
5						X		X
6							X	
7								X
8								

pode ser convertido no fatorial:

	1	3	5	7
2	X	X	X	X
4	X	X	X	X
6	X	X	X	X
8	X	X	X	X

Considerando esses tipos de dialélicos parciais, as seguintes equivalências são: fatorial 2×2 e dialélico parcial com 4 genitores; fatorial 3×3 e dialélico parcial com 6 genitores; fatorial 4×4 e dialélico parcial com 8 genitores; fatorial 5×5 e dialélico parcial com 10 genitores e assim sucessivamente.

Quanto ao meio dialélico desconexo, o índice multiefeitos associado utiliza os mesmos efeitos que o fatorial desconexo e os dois delineamentos são mais comparáveis quando se considera o mesmo número de cruzamentos por genitor

em ambos os delineamentos. Assim, no caso do número fixo de 100 famílias, os dois delineamentos podem ser caracterizados conforme a Tabela 15.6.

Tabela 15.6. Caracterização de alguns delineamentos fatoriais desconexos e meio dialélicos desconexos em termos de números de cruzamentos por conjunto, número de conjuntos e número de genitores utilizados, para um número total fixo de 100 famílias de irmãos germanos.

	Dimensão	Nº. de cruzamentos por conjuntos	Nº. de conjuntos	Nº. de genitores
Fatorial desconexo	2 x 2	4	25,00	100
Meio dialélico desconexo	3	3	33,33	100
Fatorial desconexo	3 x 3	9	11,11	67
Meio dialélico desconexo	4	6	16,67	67
Fatorial desconexo	4 x 4	16	6,25	50
Meio dialélico desconexo	5	10	10,00	50
Fatorial desconexo	5 x 5	25	4,00	40
Meio dialélico desconexo	6	15	6,67	40

Considerando o mesmo número total de: genitores utilizados, de cruzamentos gerados e de cruzamentos por genitor, verifica-se que a única diferença entre os dois delineamentos refere-se ao número de cruzamentos dentro de cada conjunto. Como os índices relativos aos dois delineamentos usam as mesmas fontes de informação, a única diferença na acurácia destes índices é decorrente do fato de que a média geral de cada conjunto advém de diferentes números de observações, e está associada a um menor número de observações no meio dialélico desconexo. Este fato conduz à menor eficiência do meio dialélico desconexo devida ao efeito do tamanho amostral finito e à correção para este efeito, proporcionada pelos índices multiefeitos. No caso do fatorial desconexo 4 x 4 e do meio dialélico com 5 genitores, a média do conjunto advém de 16 e 10 cruzamentos, respectivamente (Tabela 15.6). Assim, maior porção de variância genética não é utilizada devido à amostra finita (Resende & Higa, 1994a) no meio dialélico do que no fatorial, levando à superioridade do fatorial desconexo. Este resultado é válido para todos os conjuntos pequenos comparáveis. Para conjuntos grandes, meios dialélicos desconexos e fatoriais desconexos se equivalem. Estas inferências teóricas estão de acordo com os resultados de simulação apresentados por Van Buijtenen & Burdon (1990).

Os resultados revelaram, assim, uma vantagem ligeira do fatorial 4 x 4 sobre os demais fatoriais desconexos e sobre o meio dialélico desconexo para o ganho genético no curto prazo. Para o melhoramento no longo prazo, Kang (1979) e Kang & Namkoong (1980) concluíram que o balanceamento (isto é, a participação eqüitativa de todos os genitores) é mais importante na minimização da taxa de perda de alelos do que o tipo de delineamento. A questão do balanceamento é importante, também, no melhoramento de curto prazo, já que permite comparações mais adequadas entre os indivíduos na população de seleção e, também, maior eficiência seletiva (Van Buijtenen & Burdon, 1990). Assim, os fatoriais quadrados devem ser preferidos aos retangulares, exceto se maior ênfase for necessária a determinados genitores com maiores valores genéticos preditos. Neste último caso, os melhores genitores poderiam ser utilizados na categoria menos numerosa do fatorial.

Os resultados referentes às comparações entre os delineamentos de policruzamento (progênies de meios-irmãos), pares simples (progênies de irmãos germanos) e fatorial desconexo 4 x 4 são apresentados nas Tabelas 15.7 e 15.8 (para delineamentos com 6 plantas por parcela) e 15.9 e 15.10 (para delineamentos com 1 planta por parcela).

Tabela 15.7. Acurárias seletivas associadas aos delineamentos de policruzamento (r_{Ial}^{MI}), pares simples (r_{Ial}^{IG}) e fatorial desconexo 4 x 4 (r_{IaF_4}), bem como eficiências do delineamento de pares simples sobre o policruzamento (E_1), do fatorial sobre policruzamentos (E_2) e do fatorial sobre pares simples (E_3). Delineamento experimental de blocos ao acaso com 6 plantas por parcela e propagação sexuada.

	h_a^2*	h_g^2*	r_{Ia1}^{MI}	r_{Ia1}^{IG}	r_{IaF_4}	E_1	E_2	E_3
0,05	0,05	0,38	0,56	0,62	1,47	1,63	1,11	
	0,075	0,38	0,52	0,59	1,37	1,55	1,13	
0,10	0,10	0,46	0,63	0,66	1,37	1,43	1,05	
	0,15	0,46	0,58	0,64	1,26	1,39	1,10	
0,20	0,20	0,56	0,69	0,71	1,23	1,27	1,03	
	0,30	0,56	0,64	0,68	1,14	1,21	1,06	
0,30	0,30	0,63	0,73	0,74	1,16	1,17	1,01	
	0,45	0,63	0,68	0,71	1,08	1,13	1,04	
0,40	0,40	0,69	0,77	0,77	1,12	1,12	1,00	
	0,60	0,69	0,72	0,75	1,04	1,09	1,04	
0,50	0,50	0,75	0,81	0,81	1,08	1,08	1,00	
	0,75	0,75	0,77	0,80	1,03	1,07	1,04	

* h_a^2 e h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

Tabela 15.8. Acurárias seletivas associadas aos delineamentos de policruzamento (r_{Ig2}^{MI}), pares simples (r_{Ig2}^{IG}) e factorial desconexo 4 x 4 (r_{IgF_4}), bem como eficiências do delineamento de pares simples sobre o policruzamento (E_1), do factorial sobre policruzamentos (E_2) e do factorial sobre pares simples (E_3). Delineamento experimental de blocos ao acaso com 6 plantas por parcela e propagação assexuada.

$h_a^2 *$	$h_g^2 *$	r_{Ig2}^{MI}	r_{Ig2}^{IG}	r_{IgF_4}	E_1	E_2	E_3
0,05	0,05	0,38	0,56	0,62	1,47	1,63	1,11
	0,075	0,41	0,57	0,65	1,39	1,59	1,14
0,10	0,10	0,46	0,63	0,66	1,37	1,43	1,05
	0,15	0,52	0,64	0,71	1,23	1,37	1,11
0,20	0,20	0,56	0,69	0,71	1,23	1,27	1,03
	0,30	0,65	0,72	0,79	1,11	1,22	1,10
0,30	0,30	0,63	0,73	0,74	1,16	1,17	1,01
	0,45	0,75	0,79	0,85	1,05	1,13	1,08
0,40	0,40	0,69	0,77	0,77	1,12	1,12	1,00
	0,60	0,84	0,86	0,92	1,02	1,10	1,07
0,50	0,50	0,75	0,81	0,81	1,08	1,08	1,00
	0,75	0,93	0,95	1,00	1,02	1,08	1,05

* h_a^2 e h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

Tabela 15.9. Acurárias seletivas associadas aos delineamentos de policruzamento (r_{Ial}^{MI}), pares simples (r_{Ial}^{IG}) e factorial desconexo 4 x 4 (r_{IalF_4}), bem como eficiências do delineamento de pares simples sobre o policruzamento (E_1), do factorial sobre policruzamentos (E_2) e do factorial sobre pares simples (E_3). Delineamento experimental de blocos ao acaso com 1 planta por parcela e propagação sexuada.

$h_a^2 *$	$h_g^2 *$	r_{Ial}^{MI}	r_{Ial}^{IG}	r_{IalF_4}	E_1	E_2	E_3
0,05	0,05	0,38	0,57	0,62	1,50	1,63	1,09
	0,075	0,38	0,53	0,60	1,39	1,58	1,13
0,10	0,10	0,47	0,64	0,67	1,36	1,43	1,05
	0,15	0,47	0,59	0,64	1,26	1,36	1,09
0,20	0,20	0,57	0,71	0,71	1,25	1,25	1,00
	0,30	0,57	0,65	0,68	1,14	1,19	1,05
0,30	0,30	0,64	0,74	0,74	1,16	1,16	1,00
	0,45	0,64	0,69	0,71	1,08	1,11	1,03
0,40	0,40	0,69	0,78	0,77	1,13	1,12	0,99
	0,60	0,69	0,72	0,74	1,04	1,07	1,03
0,50	0,50	0,75	0,81	0,81	1,08	1,08	1,00
	0,75	0,75	0,76	0,77	1,01	1,03	1,01

* h_a^2 e h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

Tabela 15.10. Acurárias seletivas associadas aos delineamentos de policruzamento (r_{Ig2}^{MI}), pares simples (r_{Ig2}^{IG}) e fatorial desconexo 4 x 4 (r_{IgF_4}), bem como eficiências do delineamento de pares simples sobre o policruzamento (E_1), do fatorial sobre policruzamentos (E_2) e do fatorial sobre pares simples (E_3). Delineamento experimental de blocos ao acaso com 1 planta por parcela e propagação assexuada.

$h_a^2 *$	$h_g^2 *$	r_{Ig2}^{MI}	r_{Ig2}^{IG}	r_{IgF_4}	E_1	E_2	E_3
0,05	0,05 0,075	0,38 0,41	0,57 0,58	0,62 0,65	1,50 1,41	1,63 1,59	1,09 1,12
0,10	0,10 0,15	0,47 0,52	0,64 0,65	0,67 0,72	1,36 1,25	1,43 1,38	1,05 1,11
0,20	0,20 0,30	0,57 0,65	0,71 0,73	0,71 0,80	1,25 1,12	1,25 1,23	1,00 1,10
0,30	0,30 0,45	0,64 0,75	0,74 0,79	0,74 0,85	1,16 1,05	1,16 1,13	1,00 1,08
0,40	0,40 0,60	0,69 0,84	0,78 0,85	0,77 0,91	1,13 1,01	1,12 1,08	0,99 1,07
0,50	0,50 0,75	0,75 0,91	0,81 0,91	0,81 0,96	1,08 1,00	1,08 1,05	1,00 1,05

* h_a^2 e h_g^2 : herdabilidades individuais no sentido restrito e amplo, respectivamente.

Os resultados revelaram a superioridade do delineamento de pares simples sobre o de policruzamento em todas as situações, ou seja, em presença ou ausência de dominância, nos delineamentos com 1 e 6 plantas por parcela e nos sistemas de propagação sexuada e assexuada. As maiores eficiências foram obtidas para caracteres de mais baixa herdabilidade e em ausência de dominância ($h_g^2 = h_a^2$) (Tabelas 15.7 a 15.10).

O fatorial desconexo 4 x 4 apresentou também maior eficiência sobre o policruzamento em todas as situações e também maior eficiência que o delineamento de pares simples em quase todas as situações. A eficiência do fatorial sobre pares simples mostrou-se mais significativa (entre 5% e 14% de superioridade) para caracteres de baixa ($\leq 0,30$) herdabilidade e em presença de dominância. Considerando apenas a propagação vegetativa em presença de dominância, a superioridade do fatorial sobre pares simples ficou acima de 5% para todos os níveis de herdabilidade considerados (Tabelas 15.7 a 15.10).

A superioridade do fatorial desconexo 4×4 sobre pares simples é muito representativa, pois não implica qualquer custo adicional, já que ambos baseiam-se em cruzamentos controlados e estão associados ao mesmo número total de cruzamentos. Assim, na opção por cruzamentos controlados o fatorial desconexo 4×4 deve ser utilizado. Esta inferência pode ser estendida para programas de seleção recorrente recíproca em que se objetiva também a seleção de clones superiores dentro de progêneres associadas a cruzamentos interpopulacionais.

Na opção entre cruzamentos controlados (fatorial desconexo 4×4) e polinização livre (policruzamentos), deve prevalecer a alta eficiência (E_2) do fatorial sobretudo para caracteres de baixa herdabilidade. Para caracteres com herdabilidade menor ou igual a 0,20, a superioridade permaneceu acima de 20% em todas as situações (Tabelas 15.7 a 15.10).

Ainda na comparação entre polinização aberta e polinização controlada visando à maximização do ganho genético de curto prazo, Resende & Vencovsky (1993) compararam a seleção combinada com progêneres de autofecundação (S_1) e com progêneres de polinização aberta. Para herdabilidades no sentido restrito e amplo de 0,10 e 0,15, respectivamente, os referidos autores obtiveram uma superioridade da ordem de 45% para a seleção com S_1 . No presente trabalho, para estes mesmos níveis de herdabilidade, a superioridade do fatorial 4×4 sobre o policruzamento foi de 39%. Considerando os vários problemas (baixa produção de sementes, ausência de coincidência de florescimento entre os vários indivíduos S_1 por ocasião da recombinação, possíveis problemas experimentais) práticos potenciais associados à utilização de S_1 's, pode-se relatar que a utilização do fatorial desconexo 4×4 é uma opção muito mais segura e que apresenta ganho esperado similar à utilização de S_1 .

As Tabelas 15.7 a 15.10 permitem também comparar as eficiências (por meio das acurárias seletivas) dos sistemas de propagação assexuada e sexuada para os vários níveis de herdabilidade. Conforme esperado, constata-se que, em presença de dominância, a propagação vegetativa é sempre superior à propagação sexuada.

15.3.6. Considerações finais

- a) Dentre cruzamentos fatoriais de várias dimensões, o fatorial 4×4 praticamente conduz à maximização da eficiência seletiva.
- b) A inclusão do efeito de grupo de genitores nos índices multiefeitos é fundamental em fatoriais desconexos de pequena dimensão.
- c) Considerando o mesmo número total de genitores utilizados, o mesmo número total (na soma dos vários grupos) de cruzamentos gerados e o mesmo número de cruzamentos por genitor, o delineamento fatorial desconexo é mais eficiente que o meio dialélico desconexo.
- d) O delineamento de pares simples (irmãos germanos) mostrou-se superior ao de policruzamento (meios-irmãos), em todas as situações.

- e) Para caracteres de baixa herdabilidade ($\leq 0,30$) e em presença de dominância, a superioridade do fatorial 4×4 sobre pares simples é de 5% a 14%.
- f) Para caracteres de baixa herdabilidade, a superioridade do fatorial 4×4 sobre policruzamentos é de 20% a 63%.

15.4. Comparação de Métodos de Melhoramento

O ganho genético com seleção equivale à média dos valores genéticos preditos dos N indivíduos selecionados, e é dado por:

$$\begin{aligned} G_s &= \sum_{a_i} / N \\ &= \sum [h_a^2 (y_i - \mu)] / N \\ &= h_a^2 \sum (y_i - \mu) / N \\ &= h_a^2 (\bar{y}_s - \mu) \\ &= h_a^2 d_s \end{aligned}$$

$G_s = h_a^2 d_s$ é a expressão clássica para o cômputo do ganho genético, em que d_s refere-se ao diferencial de seleção e \bar{y}_s , à média fenotípica dos indivíduos selecionados.

O ganho genético é um indicador de eficiência na seleção e pode ser computado alternativamente por:

$$\begin{aligned} G_s &= h_a^2 d_s \\ &= k h_a^2 \sigma_y \\ &= k \sigma_a^2 / \sigma_y \\ &= k h_a \sigma_a \\ &= k r_{aa} \sigma_a \end{aligned}$$

Verifica-se nesta última expressão que h_a (raiz quadrada da herdabilidade) refere-se à acurácia (r_{aa}) seletiva da seleção massal. Assim, de maneira genérica, a expressão para o ganho genético é dada por $G_s = k r_{aa} \sigma_a$. Esta expressão do ganho é mais adequada para efeito de comparação (predição) de métodos de seleção, pois k e σ_a podem ser considerados constantes nos métodos de seleção e, portanto, a comparação pode basear-se apenas na acurácia associada aos diferentes métodos. Esta expressão permite inferir, também, que o sucesso na seleção depende, essencialmente, da utilização de acurados procedimentos de predição de valores genéticos.

Estimadores da acurácia associada aos diferentes métodos de seleção foram apresentados por Resende et al. (1995a e b). De maneira genérica, os estimadores são:

a) Acurácia e ganho associados à predição e seleção pelos efeitos aditivos

$$\hat{r}_{aa} = \frac{\hat{\sigma}_I}{\hat{\sigma}_a} = \frac{\left[\sum b_i^2 V(i) \right]^{1/2}}{\hat{\sigma}_a}; \quad G_s = K \hat{r}_{aa} \hat{\sigma}_a$$

b) Acurácia e ganhos associados à predição e seleção pelos efeitos genotípicos

$$\hat{r}_{gg} = \frac{\hat{\sigma}_I}{\hat{\sigma}_g} = \frac{\left[\sum b_i \hat{V}(i) \right]^{1/2}}{\hat{\sigma}_g}; \quad G_s = K \hat{r}_{gg} \hat{\sigma}_g, \text{ em que:}$$

$\hat{\sigma}_I$ = estimativa do desvio padrão do índice de predição dos valores genéticos.

b_i e $V(i)$ = estimativas dos coeficientes de ponderação e variância do efeito i, incluído no índice de predição de valores genéticos.

Este último ganho pode ser expresso em termos do desvio padrão genético aditivo da população por $G_s = k r_{Ig} (h_g / h_a) \sigma_a$.

Na prática, a acurácia com seleção baseada em um índice formado pela soma de várias fontes de informação é dada por:

$$r_{Ia} = \left[\sum r_{Iai}^2 \right]^{1/2} \text{ para os efeitos aditivos e}$$

$$r_{Ig} = \left[\sum r_{Iai}^2 + \sum r_{Igi}^2 \right]^{1/2} \text{ para os efeitos genotípicos, em que:}$$

$$\sum r_{Iai}^2 = \frac{\sum b_i^2 V(i)}{\sigma_a^2}.$$

$$\sum r_{Idi}^2 = \frac{\sum b_i^2 V(i)}{\sigma_d^2}.$$

r_{Id} : acurácia na predição dos efeitos de dominância.

A seguir são comparados alguns métodos de seleção e sistemas de propagação associados ao melhoramento de espécies perenes. Os estimadores da acurácia (considerando experimentos com uma planta por parcela, em que $N = nb$ = b refere-se ao número de indivíduos por família ou número de blocos) e ganhos genéticos, obtidos com base em Resende et al. (1995a e b) e Resende et al. (1996d), são dados por:

(1) Propagação sexuada dos indivíduos selecionados

(1.1) Seleção baseada em valores individuais, sem o uso da informação da família.

Acurácia	Ganho genético
$r_{\hat{aa}_1} = \left[\frac{m h_a^2}{1 + (m-1) \rho} \right]^{1/2}$	$G_{s_1} = k r_{\hat{aa}_1} \sigma_a$

(1.2) Seleção de genitores baseada no comportamento da progênie de meios-irmãos.

Acurácia	Ganho genético
$r_{\hat{aa}_2} = 0,5 \left[\frac{m N h_a^2}{1 + (m-1) \rho + (N-1) m 0,25 h_a^2} \right]^{1/2}$	$G_{s_2} = k r_{\hat{aa}_2} \sigma_a$

(1.3) Seleção baseada em valores individuais e na informação da família de meios-irmãos.

Acurácia
$r_{\hat{aa}_3} = \left[\frac{0,5625 m \frac{N-1}{N} h_a^2}{1 + (m-1) \rho - m 0,25 h_a^2} + \frac{m [1 + (N-1) 0,25]^2 h_a^2}{N [1 + (m-1) \rho + (N-1) m 0,25 h_a^2]} \right]^{1/2}$

Ganho genético

$$G_{s_3} = k r_{\hat{aa}_3} \sigma_a$$

(1.4) Seleção baseada em valores individuais e na informação da família de irmãos germanos.

Acurácia
$r_{\hat{aa}} = \left[\frac{0,25 m \frac{N-1}{N} h_a^2}{1 + (m-1) \rho - m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]} + \frac{m [1 + (N-1) 0,5]^2 h_a^2}{N [1 + (m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]]} \right]^{1/2}$

Ganho genético

$$G_{s_4} = k r_{\hat{aa}_4} \sigma_a$$

(1.5) Seleção de genitores baseada no comportamento em teste clonal.

Acurácia

$$r_{\hat{aa}_5} = \left[\frac{m N h_a^2}{1 + (m-1) \rho + (N-1) m h_g^2} \right]^{1/2}$$

Ganho genético

$$G_{s_5} = k r_{\hat{aa}_5} \sigma_a$$

(1.6) Seleção de cruzamentos biparentais, baseada no comportamento da progenie de irmãos germanos e visando ao plantio (via sementes) comercial destas (reconstituídas a cada geração).

Neste caso, a acurácia deve ser calculada em separado para os efeitos aditivos ($r_{\hat{aa}_6}$) e de dominância ($r_{\hat{dd}_6}$) e, posteriormente, obtém-se a acurácia para os efeitos genotípicos ($r_{\hat{gg}_6}$), procedendo-se da seguinte maneira:

$$\begin{aligned} r_{\hat{aa}_6} &= \left[\frac{m N 0,25 h_a^2}{1 + (m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]} \right]^{1/2}; \\ r_{\hat{dd}_6} &= \left[\frac{m N 0,0625 (h_g^2 - h_a^2)}{1 + (m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]} \right]^{1/2}; \\ r_{\hat{gg}_6} &= \left[r_{\hat{aa}_6}^2 + r_{\hat{dd}_6}^2 \right]^{1/2}; \end{aligned}$$

Ganho genético

$$G_{s_6} = k r_{\hat{gg}_6} \sigma_g = k r_{\hat{gg}_6} (h_g / h_a) \sigma_a$$

Alternativamente, o ganho genético é dado por:

$$G_s = k r_{\hat{aa}} \sigma_a + k r_{\hat{dd}} \sigma_d = k \left[r_{\hat{aa}} + r_{\hat{dd}} \left[\frac{h_a^2 - h^2}{h^2} \right]^{1/2} \right] \sigma_a$$

- (1.7) Seleção de genitores baseada no efeito da capacidade geral de combinação sob cruzamentos fatoriais ou dialélicos.

As expressões para a acurácia e ganho genético foram apresentadas por Resende (1999b). Considerando os cruzamentos fatoriais, a acurácia para a seleção dos genitores femininos é dada por:

$$r_{Ia_f} = 0,5 \left[\frac{nbm h_a^2}{1+(n-1)[0,5 h_a^2 + 0,25 (h_g^2 - h_a^2) + c^2] + (b-1) n [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)] + (m-2) nb 0,25 h_a^2} \right]^{1/2}$$

A acurácia para a seleção com base no efeito da capacidade geral de combinação do genitor masculino r_{Ia_m} pode ser obtida pela mesma fórmula, porém substituindo m por f.

- (1.8) Seleção baseada em valores individuais e nas informações das famílias de irmãos germanos e de meios-irmãos obtidas sob cruzamentos fatoriais ou dialélicos.

As expressões para a acurácia foram apresentadas por Resende (1999b) e encontram-se no tópico 15.3.3.

- (1.9) Idem (1.6), porém com seleção baseada em $(1/2) a_i + (1/2) a_j + d_{ij}$, onde a_i e d_{ij} referem-se à capacidade geral e específica de combinação, respectivamente, obtidas sob cruzamentos fatoriais ou dialélicos.

As expressões para a acurácia e ganho genético foram apresentadas por Resende (1999b).

$$r_{Ia} = \left[\frac{1}{4} (r_{Ia_f}^2 + r_{Ia_m}^2) \right]^{1/2} : \text{acurácia na predição dos efeitos aditivos.}$$

$$r_{Id} = 0,25 \left[\frac{nb (h_g^2 - h_a^2)}{1+(n-1)[0,5 h_a^2 + 0,25 (h_g^2 - h_a^2) + c^2] + (b-1) n [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)] - 2nb 0,25 h_a^2} \right]^{1/2} : \\ \text{acurácia na predição dos efeitos de dominância.}$$

$$r_{Ig} = \left[r_{Ia}^2 + r_{Id}^2 \right]^{1/2} : \text{acurácia na predição do efeito genotípico da progénie a ser plantada.}$$

O ganho genético neste caso é dado por:

$$G_s = k r_{Ig} \sigma_g = k r_{Ig} h_g / h_a \sigma_a.$$

(2) Propagação assexuada dos indivíduos selecionados

(2.1) Seleção baseada em valores individuais, sem o uso da informação da família ou de testes clonais.

Acurácia	Ganho Genético
$r_{\hat{g}g_7} = \left[\frac{m h_g^2}{1 + (m-1) \rho} \right]^{1/2}$	$G_{s_7} = k r_{\hat{g}g_7} \sigma_g$ $= k r_{\hat{g}g_7} (h_g / h_a) \sigma_a$

(2.2) Seleção de indivíduos baseada no comportamento em teste clonal.

Acurácia	Ganho genético
$r_{\hat{g}g_8} = \left[\frac{m N h_g^2}{1 + (m-1) \rho + (N-1) m h_g^2} \right]^{1/2}$	$G_{s_8} = k r_{\hat{g}g_8} \sigma_g = k r_{\hat{g}g_8} (h_g / h_a) \sigma_a$

(2.3) Seleção baseada em valores individuais e na informação da família de meios-irmãos.

Acurácia	Ganho genético
$r_{\hat{g}g_9} = \left[\frac{m \frac{N-1}{N} [0,5625 h_a^2 + (h_g^2 - h_a^2)]}{1 + (m-1) \rho - m 0,25 h_a^2} + \frac{m [(1+(N-1) 0,25)^2 h_a^2 + (h_g^2 - h_a^2)]}{N [1 + (m-1) \rho + (N-1) m 0,25 h_a^2]} \right]^{1/2}$	$G_{s_9} = k r_{\hat{g}g_9} \sigma_g = k r_{\hat{g}g_9} (h_g / h_a) \sigma_a$

(2.4) Seleção baseada em valores individuais e na informação da família de irmãos germanos.

Acurácia	Ganho genético
$r_{\hat{g}g_{10}} = \left[\frac{m \frac{N-1}{N} [0,25 h_a^2 + 0,5625 (h_g^2 - h_a^2)]}{1 + (m-1) \rho - m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]} + \frac{m [(1+(N-1) 0,5)^2 h_a^2 + (1+(N-1) 0,25)^2 (h_g^2 - h_a^2)]}{N [1 + (m-1) \rho + (N-1) m [0,5 h_a^2 + 0,25 (h_g^2 - h_a^2)]]} \right]^{1/2}$	$G_{s_{10}} = k r_{\hat{g}g_{10}} \sigma_g = k r_{\hat{g}g_{10}} (h_g / h_a) \sigma_a$

Ganho genético

$$G_{s_{10}} = k \ r_{\hat{g}g_{10}} \ \sigma_g = k \ r_{\hat{g}g_{10}} \ (h_g / h_a) \ \sigma_a$$

- (2.5) Seleção de indivíduos (clones) baseada no comportamento da progênies de meios-irmãos.

Acurácia

$$r_{\hat{g}g_{11}} = 0,5 \left[\frac{m N h_a^2}{1 + (m-1) \rho + (N-1) m 0,25 h_a^2} \right]^{1/2}$$

Ganho genético

$$G_{s_{11}} = k \ r_{\hat{g}g_{11}} \ \sigma_a$$

- (2.6) Seleção baseada em valores individuais e nas informações das famílias de irmãos germanos e de meios-irmãos obtidas sob cruzamentos fatoriais ou dialélicos

As expressões para a acurácia foram apresentadas por Resende (1999b) e encontram-se no tópico 15.3.3.

(3) Manutenção dos indivíduos selecionados no mesmo microambiente

- (3.1) Seleção pelo valor fenotípico permanente

Acurácia

$$r_{\hat{p}p_{12}} = \left[\frac{m \rho}{1 + (m-1) \rho} \right]^{1/2}$$

Ganho com seleção

$$G_{s_{12}} = k \ r_{\hat{p}p_{12}} \ \sigma_{fp} = k \ r_{\hat{p}p_{12}} \ (\rho / h_a^2)^{1/2} \ \sigma_a$$

Nas expressões apresentadas, $N = nb$ refere-se ao número total de indivíduos por progénies, $r_{\hat{p}p}$ refere-se à acurácia na predição dos efeitos fenotípicos permanentes e σ_{fp}^2 à variância destes efeitos.

Verifica-se pelas expressões (em 1.2 e 2.5) apresentadas que, tendo sido feita a seleção pelo teste de progénies, o plantio dos indivíduos selecionados, via sementes (seleção em ambos os sexos) ou propágulos clonais conduz ao mesmo ganho genético. Se a seleção foi feita pelo teste clonal (expressões em

1.5 e 2.2), o plantio por via vegetativa é h_g^2/h_a^2 vezes mais eficiente que o plantio via sementes, ou seja, só é mais eficiente se houver dominância.

Considerando ainda estas duas formas de teste (clonal e de progênies) e os dois sistemas de propagação (sexuada e assexuada), Resende et al. (1998b) concluíram que para um caráter com $h_a^2 = 0,20$:

- (i) Quando o objetivo é realizar o plantio via sementes:
 - Com $N < 30$, o teste clonal é sempre melhor que o teste de progênies.
 - Com $N = 30$ e em presença de dominância completa, as duas formas de teste se equivalem.
 - Em ausência de dominância, o teste clonal é sempre melhor que o teste de progênies.
 - Com dominância completa e $N = 100$, o teste de progênies é 14% superior.
- (ii) Quando o objetivo é realizar o plantio clonal:
 - A seleção pelo teste clonal é sempre mais vantajosa que pelo teste de progênies.
 - A seleção pelo teste clonal é tanto mais vantajosa quanto menor N e quanto maior a dominância.

Verifica-se que os sistemas de teste e propagação apresentados exploram diferentes quantidades da variabilidade fenotípica populacional. Os sistemas de propagação (1), (2) e (3) exploram, em geral, a variância aditiva (σ_a^2), genotípica (σ_g^2) e fenotípica permanente (σ_{fp}^2), respectivamente. Entretanto, algumas exceções existem: a seleção pelo teste de progênies e propagação por via vegetativa (sistema 2.5) explora apenas a variação aditiva, não havendo, na seleção, controle sobre os efeitos não aditivos; a seleção de cruzamentos biparentais (1.6 e 1.9) explora também uma fração da variação de dominância. Assim, nem sempre a propagação vegetativa explora os efeitos não aditivos e também, às vezes, a propagação por sementes permite explorar os efeitos não aditivos.

Todas as expressões de ganho apresentadas foram expressas em termos de σ_a de forma que a eficiência dos 16 métodos de seleção-propagação pode ser comparada se forem conhecidos os parâmetros h_a^2 , h_g^2 e ρ , visto que σ_a pode ser ignorado por ser constante nos vários métodos.

Considerando um caráter com os parâmetros $h_a^2 = 0,20$; $h_g^2 = 0,30$ e $\rho = 0,65$ (valores estes que referem-se às médias destes parâmetros para as espécies perenes) e a avaliação de uma população de tamanho (6.000 plantas) fixo, tem-se na Tabela 15.11 as eficiências dos diferentes métodos. Para os testes de progênies considerou-se $N = 60$ plantas (número este que permite atingir 95%

da acurácia máxima possível na seleção para os efeitos aditivos) e a avaliação de 100 famílias. Para os testes clonais, considerou-se $N = 20$ ramegos (número que permite atingir uma acurácia de 95% na seleção de clones) por clone e a avaliação de 300 clones. Adicionalmente, foram levados em conta: $m = 1$ e a seleção dos 10 melhores indivíduos em cada método, exceto no método 3.1 em que foi considerada a seleção de 50% da população e no método 1.6 em que selecionaram-se 5 cruzamentos biparentais (10 indivíduos). Nos cruzamentos fatoriais, considerou-se a avaliação de 50 genitores em arranjos 4 x 4.

Tabela 15.11. Comparação de métodos que objetivam o melhoramento de um caráter ($h_a^2 = 0,20$; $h_g^2 = 0,30$) quantitativo em espécie perene.

	Método	Acurácia	Acurácia*	Diferencial K	Eficiência**	Ganho genético***	Eficiência****
Propagação sexuada	1.1	0,45	0,45	3,23	1,00	1,45	1,00
	1.2	0,87	0,87	1,76	1,93	1,53	1,06
	1.3	0,57	0,57	3,23	1,27	1,84	1,27
	1.4	0,65	0,65	3,23	1,44	2,10	1,45
	1.5	0,77	0,77	1,76	1,71	1,36	0,94
	1.6	0,61	0,75	2,06	1,67	1,55	1,07
	1.7	0,91	0,91	1,40	2,02	1,27	0,88
	1.8	0,68	0,68	3,23	1,51	2,20	1,52
	1.9	0,76	0,93	2,06	2,07	1,92	1,32
Propagação assexuada	2.1	0,55	0,67	3,23	1,49	2,16	1,49
	2.2	0,95	1,16	2,23	2,58	2,59	1,79
	2.3	0,65	0,80	3,23	1,78	2,58	1,78
	2.4	0,73	0,89	3,23	1,98	2,87	1,98
	2.5	0,87	0,87	1,76	1,93	1,53	1,06
	2.6	0,80	0,98	3,23	2,18	3,17	2,19
	3.1	0,81	1,46	0,80	3,24	1,17	0,81

* Acurácia expressa em termos do desvio padrão genético aditivo de forma que todos os métodos podem ser comparados com base nela; ** Eficiência(razão entre acuráncias*) dos métodos expressa em relação à seleção massal para plantio via sementes; *** Ganho genético em unidades de desvio padrão genético aditivo, considerando as diferentes intensidades de seleção associadas aos vários métodos; **** Eficiência (razão entre ganhos genéticos) em relação à seleção massal para plantios via sementes, considerando as diferentes intensidades de seleção.

Pela Tabela 15.11 verifica-se que os sistemas que permitem maior acurácia* são: seleção e manutenção dos indivíduos selecionados no mesmo microambiente (3.1); seleção de clones em testes clonais (2.2); seleção de indivíduos (para clonagem) em testes de famílias obtidas sob cruzamentos fatoriais (2.6); seleção de cruzamentos biparentais a partir de cruzamentos fatoriais (1.9). Em termos de eficiência de ganho genético, os melhores sistemas são: seleção de indivíduos (para clonagem) em testes de famílias obtidas sob cruzamentos fatoriais (2.6); seleção de indivíduos (para clonagem) em testes de famílias de

irmãos germanos (2.4); seleção de clones a partir de testes clonais ou seleção de indivíduos (para clonagem) em testes de famílias de meios-irmãos (métodos 2.2 ou 2.3). Para a seleção visando à propagação sexuada, o melhor método também é aquele associado ao delineamento fatorial (1.8).

Os resultados revelam que a melhor maneira de explorar os efeitos de dominância e/ou da capacidade específica de combinação é a clonagem de indivíduos selecionados em testes de famílias de irmãos germanos (sistemas 2.6 e 2.4). Deve-se ressaltar, também, a grande eficiência do delineamento fatorial para o melhoramento de espécies perenes, tanto para o plantio via sexuada quanto assexuada. Os resultados obtidos podem também ser extrapolados para progênies interpopulacionais ou interespecíficas, bastando para isto considerar as variâncias e parâmetros interpopulacionais.

É importante relatar também o efeito da intensidade de seleção na eficiência relativa dos sistemas. Considerando a população experimental com tamanho fixo (6.000 plantas), verifica-se que a seleção dos 10 melhores indivíduos pelo sistema (2.6) conduz a um ganho genético 22% (2,19/1,79) superior à seleção dos 10 melhores clones em um teste de 300 clones (sistema 2.2).

O sistema (3.1), embora sendo o mais acurado, implica seleção no próprio campo do produtor, fato que limita a intensidade de seleção. Tal sistema seria especialmente indicado para plantios heterogêneos sob sistemas agroflorestais (Resende et al., 1996d) ou para culturas como o cacau, cujo ciclo produtivo é longo e o cacauicultor não pode aguardar até que o novo plantio comece a produzir (Dias, 1998a). Neste caso, a produtividade do plantio pode ser elevada, simplesmente, através da substituição das piores plantas por outras com maior grau de melhoramento, com base na produção média de várias safras.

15.5. Seleção de Espécies e Populações

O sucesso de qualquer programa de melhoramento genético depende da avaliação e seleção do germoplasma mais adequado às finalidades e objetivos do cultivo das plantas. De maneira geral, os recursos genéticos de cada cultivo são muito amplos e apenas uma pequena parte deles é usada nos programas de melhoramento. Por exemplo, o gênero *Eucalyptus* apresenta cerca de 700 espécies, das quais apenas 5 (menos de 1%) são utilizadas comercialmente para a produção industrial de celulose. Cotterill & Brolin (1997) demonstraram que as variações entre espécies, entre populações e dentro de populações de *Eucalyptus* são suficientes para o desenvolvimento de diferentes tipos de produtos dentro do próprio setor de papel e celulose. Este fato ressalta a importância do conhecimento dos recursos genéticos disponíveis para cada espécie ou gênero.

A etapa inicial básica de qualquer programa de melhoramento é a seleção das espécies e populações a serem trabalhadas. Esta seleção deve ser fundamentada em testes de espécies e populações, as quais devem ser avaliadas quanto aos caracteres relacionados com o produto de interesse.

A partir da avaliação experimental, a seleção deve basear-se tanto em componentes de médias quanto em componentes de variância. Idealmente, devem ser selecionados materiais genéticos com elevada média e ampla variabilidade genética, que deverá propiciar ganhos genéticos contínuos com seleção ao longo de várias gerações. A importância da média da população e sua variabilidade genética pode ser avaliada, simultaneamente, pela expressão da média da população melhorada dada por $\bar{X}_m = \bar{X}_o + k h_a \sigma_a$, em que \bar{X}_o é a média original da população, h_a é a raiz quadrada da herdabilidade no sentido restrito e σ_a é a raiz quadrada da variância genética aditiva. A seleção da população mais adequada pode ser baseada no valor calculado de \bar{X}_m .

Na seleção de populações, linhagens e espécies para programas de hibridação em cultivos anuais, têm sido usadas as técnicas de cruzamentos dialélicos e da estimativa da divergência genética multivariada entre populações (Vencovsky & Barriga, 1992; Ramalho et al., 1993; Cruz & Regazzi, 1994; Cruz, 1997). Em espécies perenes, os cruzamentos dialélicos entre várias populações podem ser proibitivos por causa das dificuldades da polinização controlada. A estimativa da divergência genética multivariada tem sido aplicada (Resende & Higa, 1991; Dias, 2001) com vistas ao agrupamento de germoplasmas e à seleção de germoplasmas contrastantes.

A seleção de espécies e populações deve considerar também as condições edáficas (tipos de solo) e climáticas (precipitação, déficit hídrico, altitude, latitude, temperatura e classificação climática) das regiões de plantio.

15.6. Seleção de Cruzamentos Dentro de Populações

A seleção de cruzamentos dentro de populações de espécies perenes deve basear-se nos valores genéticos preditos dos indivíduos selecionados, previsões estas baseadas em testes de progêniens e/ou cruzamentos dialélicos.

A estimativa de ganhos genéticos com seleção baseia-se, de maneira geral, na contribuição equitativa dos indivíduos selecionados, ou seja, em geral todos os indivíduos selecionados contribuem com a mesma proporção de gametas para a geração da população melhorada. Entretanto, no melhoramento de espécies perenes, existe a possibilidade de contribuição desigual dos indivíduos para a geração subsequente, de forma que a eficiência da seleção pode ser aumentada por meio da utilização em maior proporção dos indivíduos com os maiores valores genéticos. Assim, via de regra, no melhoramento de plantas perenes, são realizados cruzamentos preferenciais e a adoção desta modalidade de cruzamentos para geração da população melhorada implica procedimentos de seleção de cruzamentos.

A seleção de cruzamentos pode ser empregada com sucesso quando se deseja aumentar o ganho genético. Outra aplicação da contribuição desigual dos indivíduos selecionados é no aumento do tamanho efetivo populacional pa-

ra uma condição de mesmo ganho genético, por meio da utilização, em proporções variadas, de maior número de indivíduos selecionados, porém mantendo a mesma intensidade de seleção (Toro & Nieto, 1984; Resende & Resende, 1997).

Para implementação destas duas aplicações da seleção de cruzamentos, procedimentos de predição da descendência de cada cruzamento, utilizando os valores genéticos preditos dos indivíduos, têm sido desenvolvidos (Allaire, 1980; Toro & Silio, 1992). Neste contexto, podem ser selecionados “cruzamentos preferenciais positivos” e “cruzamentos corretivos”. Os cruzamentos preferenciais positivos visam maximizar a resposta para um caráter, ao passo que os cruzamentos corretivos têm a finalidade de corrigir defeitos da progênie de determinados indivíduos que apresentam altos valores genéticos para determinados caracteres, mas baixos valores genéticos para outros caracteres de interesse.

O procedimento de seleção de cruzamentos desenvolvido por Allaire (1980) objetiva selecionar da maneira mais eficiente possível um genitor masculino para cruzar com um particular genitor feminino. A seleção do cruzamento ótimo depende dos atributos dos genitores feminino e masculino. Tal procedimento baseia-se na teoria dos índices de seleção em que:

$$I = w_1 a_1 + w_2 a_2 + \dots + w_K a_K, \text{ em que:}$$

I = índice de seleção para determinado indivíduo.

w_i e a_i = valor econômico e valor genético predito do indivíduo, respectivamente, para o caráter i .

$$I_p = w_1 (1/2)(a_1^* + a'_1) + w_2 (1/2)(a_2^* + a'_2) + \dots + w_K (1/2)(a_K^* + a'_K), \text{ e}$$

I_p = índice de predição do valor econômico global da progênie.

a_i^* e a'_i = valores genéticos preditos dos genitores femininos e masculinos, respectivamente.

Os valores da expressão de I_p podem ser rearranjados de forma a expressar os índices de seleção para genitores femininos e masculinos, separadamente:

$$I_p = (1/2) \left(\sum_i w_i a_i^* \right) + (1/2) \left(\sum_i w_i a'_i \right), \\ = (1/2) I^* + (1/2) I'$$

em que : I^* e I' são os índices de seleção para genitores femininos e masculinos, respectivamente.

Assim, a seleção de cruzamentos para maximizar o valor econômico global da progênie pode basear-se somente na escolha de indivíduos com os maiores valores econômicos globais (agregado genético), preditos pelos índices de seleção para genitores femininos e masculinos. Para caracteres que exibem heterose, um parâmetro adicional que pode ser utilizado na seleção de cruzamentos é a divergência genética entre os indivíduos, conforme apresentado por Cruz & Regazzi (1994) e Resende & Rosa-Perez (1999a).

Para o aumento do tamanho efetivo (N_e) com vistas a uma condição de ganho genético fixo, Toro & Nieto (1984) mostraram que a utilização de N' indivíduos na proporção f_i cada um, dentre um total de N avaliados, pode conduzir a um aumento do N_e , desde que N' e f_i sejam obtidos de forma a minimizar $\sum_i^N f_i^2$ (isto é, maximizando o N_e , visto que o N_e é dado por $N_e = 1/\sum_i^N f_i^2$), sujeito à restrição $\sum_i^N f_i X_i = k$ (isto é, mantendo a intensidade de seleção (k) constante), onde X_i é o valor esperado de i -ésima estatística de ordem de uma distribuição normal padrão. E esta minimização pode ser realizada empregando-se técnicas de programação linear. O exemplo de Toro & Nieto (1984) é apresentado na Tabela 15.12.

Tabela 15.12. Número de genitores selecionados (N') dentre 20 avaliados e suas proporções de contribuição (f_i), bem como o tamanho efetivo (N_e) associado, para uma intensidade de seleção fixa $k = 1,214$.

	$N' = 5$	$N' = 6$	$N' = 7$	$N' = 8$	$N' = 9$	$N' = 10$
f_1	0,20	0,24	0,26	0,27	0,27	0,27
f_2	0,20	0,19	0,20	0,20	0,20	0,20
f_3	0,20	0,17	0,16	0,16	0,15	0,15
f_4	0,20	0,15	0,13	0,12	0,12	0,12
f_5	0,20	0,13	0,11	0,10	0,10	0,10
f_6	-	0,12	0,09	0,07	0,07	0,07
f_7	-	-	0,07	0,05	0,05	0,05
f_8	-	-	-	0,03	0,03	0,03
f_9	-	-	-	-	0,01	0,01
f_{10}	-	-	-	-	-	0,00
N_e	5,00	5,67	5,71	5,84	5,94	5,94

Verifica-se que o N_e é, praticamente, maximizado (aumentando de 5 para 5,94) com a seleção dos 9 melhores indivíduos em diferentes proporções. O ganho genético é mantido constante com a seleção de 5 (em iguais proporções), 6, 7, 8 ou 9 indivíduos (Tabela 15.12).

Este resultado pode, por exemplo, ser extrapolado para a seleção clonal. Em vez de usar apenas os cinco melhores clones dentre 20 avaliados, podem ser usados nove clones em diferentes proporções, mantendo-se o mesmo ganho genético e aumentando o N_e em 20%.

15.7. Cruzamentos Preferenciais

Os dois fatores que promovem o aparecimento de cruzamentos não aleatórios são a endogamia e os cruzamentos preferenciais. Estes últimos ocorrem quando a escolha dos indivíduos a serem cruzados é baseada no fenótipo. São

ditos preferenciais positivos – PAM – quando indivíduos semelhantes são cruzados e preferenciais negativos – NAM – quando indivíduos fenotipicamente distintos são cruzados. Estas denominações também são válidas quando a escolha dos cruzamentos é baseada nos valores genéticos dos indivíduos, tal como é feito no melhoramento genético. A seleção de cruzamentos implica o delineamento PAM, o qual, via de regra, é desbalanceado. Os cruzamentos preferenciais, embora vantajosos em termos de ganho genético no curto prazo, comprometem mais o tamanho efetivo (N_e), quando comparados aos cruzamentos aleatórios. Para um N_e fixo, os cruzamentos aleatórios conduzem a maiores ganhos genéticos no longo prazo (King & Johnson, 1993).

É importante ressaltar que, mesmo com cruzamentos não preferenciais (aleatórios), alguns pares de cruzamento são fenotipicamente similares ou dissimilares. Assim, cruzamentos preferenciais referem-se aos casos em que os parceiros são mais fenotipicamente semelhantes ou dessemelhantes do que seria esperado probabilisticamente em uma população de cruzamentos ao acaso.

O efeito do PAM é o aumento das freqüências dos genótipos homozigotos e a redução das freqüências dos genótipos heterozigotos na progénie, permanecendo inalteradas as freqüências alélicas. Neste sentido, o PAM conduz a efeitos similares ao da endogamia. Entretanto, o PAM afeta apenas alguns locos (associados ao caráter em que a escolha foi baseada), ao passo que a endogamia afeta todos os locos, provocando maior efeito geral na homozigose de todo o conjunto gênico da população. O PAM pode produzir, porém, maior efeito nestes locos específicos do que produziria a endogamia.

Os efeitos do NAM são, em geral, opostos ao do PAM, pois aumentam os heterozigotos e reduzem os homozigotos na progénie. Outra diferença é que o NAM provoca alterações nas freqüências alélicas já que os cruzamentos entre dessemelhantes fazem com que os fenótipos mais raros tenham maior sucesso reprodutivo do que os fenótipos mais comuns. Assim, alelos mais raros são favorecidos e as freqüências alélicas mudam para níveis intermediários até que os fenótipos estejam em igual freqüência.

Os PAM são usados rotineiramente no melhoramento com o propósito de aumentar o ganho genético e conduzem a resultados satisfatórios quando a herdabilidade do caráter sob seleção é superior a 30% (Mahalovich & Bridgewater, 1989). Os NAM são, por vezes, utilizados associados aos “cruzamentos corretivos”, ou seja, cruzamentos entre indivíduos com altos valores genéticos para caracteres diferentes, visando à obtenção de indivíduos com expressão fenotípica satisfatória para ambos os caracteres. Nestes casos, pode ocorrer (se a correlação entre os caracteres for negativa) que os indivíduos sejam contrastantes para cada um dos caracteres. Os cruzamentos preferenciais negativos podem ser utilizados, também, quando se deseja a exploração da heterose.

A situação mais comum, entretanto, é que os indivíduos com maiores valores genéticos para um determinado caráter sejam cruzados mais vezes que os demais, contribuindo, assim, com maior número de descendentes (cruzamentos positivos). Este caso pode ser enquadrado como um tipo especial de PAM em

que apenas a fração superior dos fenótipos (ou genótipos) semelhantes contribui com descendentes em maior número. Assim, é importante conhecer como os cruzamentos preferenciais afetam a estimativa dos componentes da variação genética e dos parâmetros genéticos.

A magnitude dos cruzamentos preferenciais pode ser avaliada por meio da correlação entre valores fenotípicos (r) e valores genéticos (r_a) dos indivíduos cruzados. A correlação r_a causa aumento da variância aditiva e, consequentemente, da herdabilidade. Falconer (1989) apresenta alguns aspectos referentes aos efeitos dos cruzamentos preferenciais: (i) os efeitos completos não são imediatos, portanto, são necessárias algumas gerações de cruzamentos aleatórios para atingir um estado de equilíbrio (os efeitos completos não serão considerados neste texto); (ii) os efeitos são dependentes do número de locos que afeta um caráter e deve ser considerado grande número deles; (iii) os efeitos na variância de dominância são pequenos e podem ser negligenciados.

A seguir são mostradas as relações entre r e r_a referentes a dois (1 e 2) indivíduos (com valores fenotípicos y_1 e y_2 e valores genéticos a_1 e a_2) que se cruzam:

$$r = \frac{COV(y_1, y_2)}{\sigma_y^2} \Rightarrow COV(y_1, y_2) = r \sigma_y^2$$

$$r_a = \frac{COV(a_1, a_2)}{\sigma_a^2} \Rightarrow COV(a_1, a_2) = r_a \sigma_a^2$$

$$r_a = \frac{COV(h_a^2 y_1, h_a^2 y_2)}{\sigma_a^2} = \frac{h_a^4 COV(y_1, y_2)}{\sigma_a^2} = r h_a^2$$

Considerando uma geração de cruzamentos preferenciais, são os seguintes os estimadores para os componentes de variância e coeficientes de herdabilidade:

$$\sigma_{al}^2 = \sigma_a^2 (1 + 0,5 r_a)$$

$$\sigma_{y1}^2 = \sigma_y^2 (1 + 0,5 r_a h_a^2)$$

$$h_{al}^2 = h_a^2 \left[\frac{1 + 0,5 r_a}{1 + 0,5 r_a h_a^2} \right]$$

Covariância média dos pais-filhos: $0,5 \sigma_a^2(1+r)$.

Covariância pai-filhos: $0,5 \sigma_a^2(1+r)$.

Covariância de irmãos germanos: $0,5 \sigma_a^2(1+r_a)$.

Regressão média dos pais-filhos: h_a^2 .

Regressão pai-filhos: $0,5 h_a^2(1+r)$.

Correlação intraclasse entre irmãos germanos (desconsiderando os efeitos não aditivos e de ambiente comum): $\frac{0,5 \sigma_a^2(1+r_a)}{\sigma_y^2(1+0,5 r_a h_a^2)} = \frac{0,5 h_a^2(1+r_a)}{1+0,5 r_a h_a^2}$.

Verifica-se, então, aumento na herdabilidade dado pelo fator $(1+0,5 r_a)/(1+0,5 r_a h_a^2)$. Assim, cruzamento preferencial é útil como um fator que aumenta a precisão (acurácia) na avaliação de indivíduos (embora aumente a endogamia para o caráter em questão).

Quanto ao processo de estimação da herdabilidade, verifica-se que a estimação via regressão filhos sobre a média dos pais não necessita correção alguma, ao passo que a estimação via regressão pais-filhos e a correlação intraclass entre irmãos germanos demandam correção em termos de r_e e r_a .

É importante ressaltar que os cruzamentos preferenciais resultam em aumento apenas da variância genética aditiva entre famílias de irmãos germanos, permanecendo constante a variância genética dentro de progêneres, a qual é devida apenas à segregação mendeliana. Esta afirmação é válida desde que o número de locos segregantes não seja muito pequeno.

15.8. Delineamento de Cruzamento com Ênfase nos Melhores Indivíduos

Para o melhoramento genético intrapopulacional de curto prazo, deve ser dada ênfase aos cruzamentos entre os indivíduos com os maiores valores genéticos, como forma de aumentar a probabilidade de obtenção de indivíduos excepcionalmente superiores nas progêniés. Considerando os 10 melhores genitores de uma população, ordenados de acordo com os seus valores genéticos preditos, o delineamento em V, apresentado a seguir permite definir os cruzamentos a serem realizados, quando o número total de cruzamentos a ser realizado é fixado em 25.

Neste caso, apenas os melhores cruzamentos são realizados e os números de cruzamentos por genitor são:

Genitor	N. de Cruzamentos	Genitor	N. de Cruzamentos
1	9	6	5
2	8	7	4
3	7	8	3
4	6	9	2
5	5	10	1

No caso do melhoramento interpopulacional (ou melhoramento intrapopulacional em espécies dióicas), os 25 melhores cruzamentos são dados por um delineamento fatorial 5×5 , envolvendo os 5 melhores genitores de cada população.

Capítulo 16

**Melhoramento de Híbridos
Interpopulacionais e Seleção
Recorrente em Espécies Perenes**

16.1. Introdução

O aspecto perene de uma espécie conduz a um impacto direto nos esquemas de melhoramento, os quais devem ser capazes de produzir resultados práticos dentro de um período de tempo relativamente curto. Dessa forma, cultivares melhoradas devem ser criadas durante os diferentes estágios dos melhoramentos populacional e interpopulacional propriamente ditos.

A maneira mais adequada de conciliar a rápida obtenção de cultivares melhoradas e o melhoramento, no longo prazo, das espécies perenes é a adoção de esquemas de seleção recorrente. Em termos genéricos, as duas principais formas de seleção recorrente são a seleção recorrente intrapopulacional –SRI – e a seleção recorrente recíproca – SRR – (Gallais, 1978; 1989). A SRI capitaliza os efeitos aditivos dos alelos e, portanto, é uma ferramenta muito poderosa para o melhoramento populacional (seleção para a capacidade geral de combinação). A SRR enfatiza o melhoramento do híbrido interpopulacional, por meio do melhoramento do valor genético aditivo e também da heterose. Os aspectos relacionados aos métodos de seleção e delineamentos de cruzamento para o melhoramento intrapopulacional já foram discutidos em tópicos anteriores. A seguir serão abordados aspectos referentes ao melhoramento de híbridos interpopulacionais.

A heterose ou vigor híbrido está associada à capacidade específica de combinação – CEC – entre os genitores, sendo função da diversidade genética entre os progenitores e da dominância alélica nos locos que controlam o caráter de interesse. Dessa forma, um fator primordial na opção pela seleção para a capacidade geral de combinação (seleção recorrente intrapopulacional para os efeitos aditivos) ou pela seleção para capacidade específica de combinação (seleção recorrente recíproca enfatizando o híbrido interpopulacional) é o conhecimento da base genética do caráter de interesse, sobretudo quanto ao parâmetro grau médio de dominância. Determinados caracteres não exibem dominância e, portanto, para o melhoramento destes, não se justifica a seleção para a CEC.

Outra modalidade de seleção muito empregada em espécies perenes é a seleção clonal, a qual deve estar relacionada à seleção para a CEC, constituindo, muitas vezes, o limite máximo da seleção para a CEC, por meio da clonagem dos melhores indivíduos dentro dos melhores cruzamentos específicos.

16.2. Estratégias de Melhoramento

A heterose tem sido verificada para a maioria dos caracteres produtivos (herdabilidades baixas) em espécies vegetais perenes (Baudouim et al., 1997). O melhoramento da heterose ou do próprio híbrido ou cruzamento interpopulacional apresenta peculiaridades em relação ao melhoramento intrapopulacional, fato que demanda uma abordagem específica deste tema em melhoramento genético.

Para o melhoramento do cruzamento entre duas populações, duas estratégias são competidoras: a seleção recorrente intrapopulacional e a seleção recorrente recíproca. Em termos puramente genéticos, a SRR apresenta ganhos genéticos (por ciclo de seleção) superiores à SRI quando (Wricke & Weber, 1986; Gallais, 1989): (i) existe sobredominância; (ii) a soma das freqüências alélicas (p_1 e p_2) das duas populações é menor que 1 (neste caso, em qualquer nível de dominância, até mesmo na ausência). A SRI apresenta ganhos superiores à SRR em ausência de sobredominância e, simultaneamente, quando $p_1 + p_2 > 1$. Estes efeitos podem ser melhor visualizados por meio da Tabela 16.1, obtida de Wricke & Weber (1986).

Tabela 16.1*. Comparação entre a seleção recorrente recíproca – SRR – e a seleção recorrente intrapopulacional – SRI - dentro de duas populações, em termos de eficiência para o melhoramento do híbrido interpopulacional: graus de dominância mínimos (em função das freqüências alélicas p_1 e p_2 , das duas populações) para que a SRR seja superior.

$p_1 \backslash p_2$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,1	-	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
0,2		-	0,0	0,0	0,0	0,0	0,0	0,0	0,38
0,3			-	0,0	0,0	0,0	0,0	0,26	0,59
0,4				-	0,0	0,0	0,22	0,45	0,71
0,5					-	0,20	0,40	0,60	0,80
0,6						-	0,56	0,71	0,86
0,7							-	0,81	0,91
0,8								-	0,95
0,9									-

* Adaptada de Wricke & Weber (1986).

Verifica-se pela Tabela 16.1 que quando as freqüências alélicas das duas populações sob melhoramento são altas ($p_1 + p_2 > 1$), a SSR só será superior à SRI para graus de dominância médios a altos. Atribuindo-se um grau médio de dominância em torno de 0,6 (valor este relatado para o caráter volume de madeira em *Eucalyptus* por Rezende & Resende, 2000), verifica-se que a SSR não é vantajosa em todas as situações, mas é na maioria, sobretudo para populações pouco ou moderadamente melhoradas. O valor 0,6 foi obtido pela relação $[2\sigma_d^2/\sigma_a^2]^{1/2}$, supondo freqüência alélica média de 0,5 (populações com nível intermediário de melhoramento).

Ainda com referência ao grau de dominância, este pode ainda ser inferido com base na igualdade $d_g^2 = \sigma_d^2/(\sigma_a^2 + \sigma_d^2) = 2pqd^2/[2pqd^2 + \{1+d(p-q)\}^2]$, apresentada por Falconer (1989), em que p e q são as freqüências alélicas dos alelos favoráveis e desfavoráveis, respectivamente, e d e a (considerado como 1, con-

forme Wei & Van der Werf, 1993) são o valor genotípico do heterozigoto e diferença entre os homozigotos, respectivamente. Nestas condições, d representa o próprio grau de dominância e pode ser verificado que para $d = 0,5$, d_g^2 atinge o seu máximo de 0,14 em $p = 0,75$ e que para $d = 0,8$, tem-se que o máximo de d_g^2 é 0,44. O valor de d_g^2 encontrado no trabalho de Rezende & Resende (2000) foi de $0,05/0,30 = 0,167$, fato que indica, provavelmente, que o grau médio de dominância seja um pouco superior a $d = 0,5$ e, portanto, associado à dominância parcial.

Quanto ao melhoramento (pela SRR) da heterose propriamente dita, tem-se que (Wricke & Weber, 1986):

- Heterose inicial = $(p_1 - p_2)^2 \cdot d$.
- Heterose após ciclos de melhoramento = $[(p_1 + \Delta p_1) - (p_2 + \Delta p_2)]^2$, em que Δp_1 e Δp_2 são os acréscimos nas freqüências alélicas das populações 1 e 2, em função da seleção. Assim, para $p_1 > p_2$, a heterose será aumentada após a seleção se $\Delta p_1 - \Delta p_2$ for positivo. E isto acontecerá se $p_1 + p_2$ for menor que 1 ou se existir sobredominância. Para $p_1 + p_2$ maior que 1, os graus de dominância mínimos para permitir o melhoramento da heterose são apresentados na Tabela 16.1. Em suma, a superioridade da SRR advém do melhoramento da heterose, e esse melhoramento pode ser conseguido mesmo com dominância parcial.

Entretanto, na comparação entre SRR e SRI para o melhoramento de espécies perenes, outras questões de ordem prática devem ser consideradas, tais como: o maior custo, o maior intervalo de gerações, a menor intensidade de seleção e o menor tamanho efetivo populacional associados à SRR. Em termos de ganho genético por unidade de tempo, as duas estratégias podem ser comparadas (com vistas ao melhoramento do híbrido interpopulacional) pela expressão a seguir, que fornece a eficiência da SRI sobre a SRR:

$$E_f = \frac{G_{SRI}}{G_{SRR}} = \frac{r_{G_{IR}} \cdot h_I \cdot k_I \cdot t_R}{h_R \cdot k_R \cdot t_I}, \text{ em que:}$$

h_I^2 e h_R^2 : herdabilidades intrapopulacional e interpopulacional (recíproca), respectivamente.

k_I e k_R : diferenciais de seleção associados à seleção intrapopulacional e recíproca, respectivamente.

t_I e t_R : intervalos de geração para a seleção intrapopulacional e recíproca, respectivamente.

$r_{G_{IR}}$: correlação genética entre progêneres intrapopulacionais e interpopulacionais.

Em geral, pode-se inferir que $k_I > k_R$ (pois na SRI, geralmente, trabalha-se com populações maiores) e que $t_R > t_I$, fatos que contribuem para maior eficiência da SRI. O valor de $r_{G_{IR}} h_I / h_R$ é totalmente dependente das propriedades genéticas do caráter nas populações específicas. Considerando $h_I/h_R \approx 1$, verifica-se que a eficiência depende, sobretudo, da correlação genética entre progêneres intrapopulacionais e interpopulacionais. Esta correlação pode ser vista como uma correlação entre a capacidade geral de combinação dos genitores dentro das populações com a capacidade geral de hibridação (Gallais, 1989) destes genitores. Assim, torna-se prioritário na comparação entre SRR e SRI, quantificar a magnitude da correlação entre capacidade geral de combinação e capacidade geral de hibridação. Para populações não divergentes ou pouco divergentes, é provável que essa correlação tenda a 1 e a SRI suplante a SRR em termos de ganho por unidade de tempo (neste caso, os ganhos por ciclo, considerando iguais intensidades de seleção, deverão ser iguais para SRI e SRR). Segundo Souza Junior (1999), para populações divergentes, as correlações entre progêneres intrapopulacionais e interpopulacionais obtidas de uma mesma planta deverão ser altas para uma população e baixas para a outra população. Neste caso, a SRR poderá conduzir a melhores resultados.

Shelbourne (1993) comparou várias estratégias de melhoramento de híbridos interespecíficos e concluiu que a SRI propicia maiores ganhos genéticos por ano e é mais robusta que a SRR em termos de tamanho efetivo populacional. Apesar desta aparente vantagem da SRI, a SRR provavelmente propicia a identificação de genótipos excepcionalmente produtivos antes que a SRI, embora no longo prazo a SRI possa suplantar a SRR. Desta forma, as duas estratégias podem ser utilizadas simultaneamente, sobretudo nos primeiros ciclos de seleção no qual a realização de cruzamentos interespecíficos propiciaria: a estimativa da capacidade geral de hibridação dos genitores; a estimativa da correlação entre capacidade geral de combinação e capacidade geral de hibridação; a identificação precoce de híbridos superiores através do direcionamento dos cruzamentos entre os genitores com maiores capacidades gerais de hibridação (Resende & Higa, 1990).

Recentemente, Souza Junior (1993; 1999) mostrou que a SRR e SRI são métodos incapazes de melhorar adequadamente os híbridos interpopulacionais e as duas populações, simultaneamente. Assim, tal autor propôs um esquema modificado denominado THS (*test-cross half-sib selection*), o qual avalia progêneres interpopulacionais para uma (a mais produtiva) população e progêneres intrapopulacionais para a outra (a menos produtiva) população. Tal método, quando aplicado em populações divergentes, conduz a cerca de 7% e 25% a mais de ganho genético, por ciclo, no híbrido interpopulacional, do que a SRR e SRI, respectivamente, para caracteres com graus médios de dominância (d/a) entre 0,75 e 1,0. Para o melhoramento das populações per se, o procedimento THS é tão eficiente quanto a SRR para a população mais produtiva e tão eficiente quanto a SRI para a população menos produtiva. É importante ressaltar que para as condições citadas anteriormente (populações divergentes e nível de

dominância entre 0,75 e 1,0) a SRR é entre 11% e 20% mais eficiente que a SRI para o melhoramento do híbrido interpopulacional, cerca de 20% mais eficiente que a SRI para o melhoramento da população mais produtiva e entre 25% e 30% menos eficiente que a SRI para o melhoramento da população menos produtiva (Souza Junior, 1999). É importante ressaltar que os resultados relatados referem-se a ganhos por ciclo de seleção e não por unidade de tempo e que foram consideradas iguais intensidades de seleção para os diferentes esquemas (SRR, SRI e THS).

Os resultados relatados por Souza Junior (1993; 1999) não são prontamente extrapoláveis para espécies perenes, pois nos esquemas que envolvem a SRI foi considerada apenas a seleção de genitores com base no comportamento de suas progênies. Em espécies perenes, a seleção intrapopulacional é geralmente baseada em índice que inclui as informações do indivíduo e da média de sua família, com recombinação dos próprios indivíduos avaliados. Isto conduz a maior intensidade de seleção e também a menor acurácia em relação à seleção de genitores. Considerando um balanço entre a menor acurácia e a maior intensidade de seleção, tem-se que a seleção pelo índice conduz a maior ganho que a seleção de genitores. Dessa forma, o esquema de Souza Júnior parece adequado também para espécies perenes.

16.3. Comparação entre Procedimentos de Seleção Recorrente Recíproca

Na opção pela SRR, dois tipos de progênies interpopulacionais podem ser prontamente empregados para espécies perenes: meios-irmãos – MI (usando mistura de pólen de pelo menos 4 ou 5 testadores da população recíproca) e irmãos germanos – IG. As expressões do ganho genético no híbrido interpopulacional, com seleção baseada nas médias das progênies e recombinação dos genitores são dadas por (Resende & Higa, 1990):

$$G_{s_1} = k_1 (1/4) \sigma_{a_{12}}^2 / \sigma_{F_{12}} + k_1 (1/4) \sigma_{a_{21}}^2 / \sigma_{F_{21}}, \text{ para progênies de meios-irmãos}$$

$$G_{s_2} = k_2 (1/4) \sigma_{a_{(12)}}^2 / \sigma_{F_{(12)}} + k_2 (1/4) \sigma_{a_{(12)}}^2 / \sigma_{F_{(12)}}, \text{ para progênies de irmãos germanos,}$$

em que:

$\sigma_{a_{12}}^2$ e $\sigma_{a_{21}}^2$: variâncias genéticas aditivas interpopulacionais que surgem quando as populações 2 e 1 são utilizadas como doadoras de pólen nos cruzamentos, respectivamente.

$\sigma_{F_{12}}$ e $\sigma_{F_{21}}$: desvios padrões fenotípicos entre médias de famílias de meios-irmãos interpopulacionais, quando as populações 2 e 1 são utilizadas como doadoras de pólen nos cruzamentos, respectivamente.

$\sigma_{a_{(12)}}^2$ e $\sigma_{F_{(12)}}^2$: variâncias genética aditiva interpopulacional e fenotípica entre médias de famílias de irmãos germanos interpopulacionais, respectivamente.

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

k_1 e k_2 : diferenciais de seleção padronizados associados à seleção de genitores nos esquemas com meios-irmãos e irmãos germanos, respectivamente.

Considerando iguais (k_1 e k_2) intensidades de seleção, e que $\sigma_{F_{(12)}}^2 = (1/2) \sigma_{a_{(12)}}^2 + (1/4) \sigma_{d_{(12)}}^2 + \text{variância residual média}$ e $\sigma_{F_{12}}^2 = (1/4) \sigma_{a_{12}}^2 + \text{variância ambiental média}$, verifica-se que o esquema empregando meios-irmãos é mais eficiente.

Expressando a herdabilidade da média de progêneres (h_m^2) como função da herdabilidade individual e o ganho genético em função de $G_s = k r_{Ia} \sigma_{a_{(12)}}$, onde r_{Ia} é a acurácia seletiva e supondo $\sigma_{a_{(12)}}^2 = (\sigma_{a_{21}}^2 + \sigma_{a_{12}}^2)/2$, tem-se:

Para a SRR com famílias de meios-irmãos:

$$h_m^2 = \frac{b(1/4) h_{(12)}^2}{1+(b-1)(1/4) h_{(12)}^2}, \text{ para experimentos com uma planta por parcela.}$$

$$h_m^2 = \frac{nb(1/4) h_{(12)}^2}{1+(n-1)[(1/4) h_{(12)}^2 + c^2] + (b-1)n(1/4) h_{(12)}^2}, \text{ para experimentos com } n \text{ plantas por parcela.}$$

$$r_{Ia_M} = \left[\frac{b(1/16) h_{(12)}^2}{1+(b-1)(1/4) h_{(12)}^2} \right]^{1/2}, \text{ para experimentos com uma planta por parcela.}$$

$$r_{Ia_M} = \left[\frac{nb(1/16) h_{(12)}^2}{1+(n-1)[(1/4) h_{(12)}^2 + c^2] + (b-1)n(1/4) h_{(12)}^2} \right]^{1/2}, \text{ para experimentos com } n \text{ plantas por parcela.}$$

Para a SRR com famílias de irmãos germanos:

$$h_m^2 = \frac{b(1/4) h_{(12)}^2}{1+(b-1)[(1/2) h_{(12)}^2 + (1/4)(h_{g(12)}^2 - h_{(12)}^2)]}, \text{ para experimentos com uma planta por parcela.}$$

$$h_m^2 = \frac{nb(1/4) h_{(12)}^2}{1+(n-1)[(1/2) h_{(12)}^2 + (1/4)(h_{g(12)}^2 - h_{(12)}^2) + c^2] + (b-1)n[(1/2) h_{(12)}^2 + (1/4)(h_{g(12)}^2 - h_{(12)}^2)]}, \text{ para experimentos com } n \text{ plantas por parcela.}$$

$$r_{Ig_G} = \left[\frac{b(1/16) h_{(12)}^2}{1+(b-1)[(1/2) h_{(12)}^2 + (1/4)(h_{g(12)}^2 - h_{(12)}^2)]} \right]^{1/2}, \text{ para experimentos com uma planta por parcela.}$$

$$r_{Ia_{IG}} = \left[\frac{nb (1/16) h_{(12)}^2}{1 + (n-1) [(1/2) h_{(12)}^2 + (1/4) (h_{g(12)}^2 - h_{(12)}^2) + c^2] + (b-1) n [(1/2) h_{(12)}^2 + (1/4) (h_{g(12)}^2 - h_{(12)}^2)]} \right]^{1/2}, \text{ para experimentos com } n \text{ plantas por parcela.}$$

Nas expressões apresentadas, são as seguintes as definições:

b : número de repetições.

$h_{(12)}^2$ e $h_{g(12)}^2$: herdabilidades individuais interpopulacionais nos sentidos restrito e amplo, respectivamente.

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

Considerando os parâmetros $h_{(12)}^2 = 0,25$, $h_{g(12)}^2 = 0,30$ e $c^2 = 0,10$, os quais foram obtidos (Rezende & Resende, 2000) a partir de populações de *E. grandis* e *E. urophylla* têm-se as seguintes acurárias e eficiências (E):

$r_{Ia_{MI}} = 0,4472$, $r_{Ia_{IG}} = 0,3208$, $E = r_{Ia_{MI}} / r_{Ia_{IG}} = 1,39$ para experimentos com uma planta por parcela e $b = 60$ blocos.

$r_{Ia_{MI}} = 0,4251$, $r_{Ia_{IG}} = 0,3123$, $E = r_{Ia_{MI}} / r_{Ia_{IG}} = 1,36$ para experimentos com $n = 6$ plantas por parcela e $b = 10$ blocos.

Verifica-se então que, para a mesma intensidade de seleção, a SRR com meios-irmãos é pelo menos 36% mais eficiente que a SRR com progêneres de irmãos germanos. É importante destacar que esquemas mais eficientes que ambos são os cruzamentos fatoriais desconexos, os dialélicos parciais e os dialélicos parciais circulantes, os quais utilizam simultaneamente meios-irmãos e irmãos germanos; entretanto, tais esquemas limitam o número de genitores a serem utilizados, quando se fixa o número total de cruzamentos a ser realizado e avaliado. Porém, para atingir a mesma intensidade de seleção, a SRR com MI necessita avaliar o dobro do número de progêneres. Considerando a avaliação de um mesmo número de progêneres na SRR com MI e na SRR com IG, tem-se na Tabela 16.2 as eficiências da seleção com MI em relação à seleção com IG, de acordo com a proporção selecionada.

Tabela 16.2. Eficiência (Ef) da seleção recorrente recíproca com meios-irmãos (MI) sobre a seleção recorrente recíproca com irmãos germanos (IG) em função das proporções selecionadas, as quais geram os diferenciais padronizados (k_1 e k_2) para MI e IG, respectivamente. Experimentação com uma planta por parcela.

Proporção selecionada						
MI	IG	K_1	K_2	$k_1 r_{Ia_{MI}} = G_{smi}$	$k_2 r_{Ia_{IG}} = G_{sig}$	$Ef = G_{smi} / G_{sig}$
0,50	0,25	0,80	1,27	0,3578	0,4074	0,8783
0,30	0,15	1,16	1,55	0,5188	0,4972	1,0434
0,20	0,10	1,40	1,76	0,6261	0,5646	1,1089
0,10	0,05	1,76	2,06	0,7871	0,6608	1,1911

$r_{Ia_{MI}}$ e $r_{Ia_{IG}}$: acurárias seletivas associadas à seleção com meios-irmãos e irmãos germanos, respectivamente.

Verifica-se assim que, para proporções de seleção igual ou inferior a 30%, a SRR com MI continua sendo superior à SRR com IG, mesmo quando se avalia metade do número de genitores em cada população pela SRR com MI.

Para a presente situação, considerando a avaliação de 200 genitores em cada população e, a seleção de 30 genitores, ocorrem as seguintes possibilidades

a) SRR com IG

- número de cruzamentos controlados = 200
- número de progênies para avaliação = 200
- intensidade de seleção: 15%
- ganho genético em unidades de $\sigma_{a(12)} = 0,4972$

b) SRR com MI

- número de cruzamentos controlados = 400
- número de progênies para avaliação = 400
- intensidade de seleção: 15%
- ganho genético em unidades de $\sigma_{a(12)} = 0,6932$

c) THS

- número de cruzamentos controlados = 200
- número de progênies para avaliação = 400
- intensidade de seleção: 15%
- ganho genético em unidades de $\sigma_{a(12)}$, considerando a eficiência de 7% (Souza Junior, 1999) sobre a SRR com MI = 0,7417

Considerando a avaliação de 100 genitores em cada população e a seleção de 30 genitores, têm-se:

d) SRR com MI

- número de cruzamentos controlados = 200
- número de progênies para avaliação = 200
- intensidade de seleção: 30%
- ganho genético em unidades de $\sigma_{a(12)} = 0,5188$

e) THS

- número de cruzamentos controlados = 100
- número de progênies para avaliação = 200
- intensidade de seleção: 30%
- ganho genético em unidades de $\sigma_{a(12)} = 0,5551$

De todas estas situações, verifica-se que aquela que propicia menor ganho é a situação (a). As eficiências de ganho genético em relação a (a) são: (b)/(a) = 1,39; (c)/(a) = 1,49; (d)/(a) = 1,04; (e)/(a) = 1,11.

Na opção por recombinar 50 genitores, os ganhos genéticos em (a), (b), (c), (d) e (e) equivalem a 0,4074; 0,5679; 0,6067; 0,3578 e 0,3828, respectivamente, e as eficiências equivalem a (b)/(a) = 1,39; (c)/(a) = 1,49; (d)/(a) = 0,8783 e (e)/(a) = 0,9396.

Considerando as dificuldades e os custos da realização de cruzamentos controlados e da avaliação de um maior número de progênies, cabe ao melhorista decidir pela melhor opção. De maneira geral, o esquema THS parece mais favorável por envolver a realização de um menor número de cruzamentos controlados, sem comprometer o ganho genético. É importante destacar que o esquema THS é o melhor também para o melhoramento das duas populações per se.

O ganho genético pela seleção com o esquema THS é dado por $G_{s_3} = k_3 (1/4) \sigma_{a_{12}}^2 / \sigma_{F_{12}} + k_3 (1/4) \sigma_{a_2 a_1} / \sigma_{F_{22}}$, considerando 1 como a população mais produtiva e 2 como a menos produtiva.

$\sigma_{a_2 a_1}$: refere-se à covariância genética aditiva entre genótipos intrapopulacionais da população 2 e interpopulacionais, sendo que $\sigma_{a_2 a_1}$ e $\sigma_{a_{12}}^2$ são dados por Souza Junior (1999):

$$\sigma_{a_2 a_1} = \sigma_{a_{22}}^2 + 2Cov_{(a_2 \tau_{21})}$$

$$\sigma_{a_{12}}^2 = \sigma_{a_{11}}^2 + \sigma_{\tau_{12}}^2 + 4Cov_{(a_1 \tau_{12})}, \text{ em que:}$$

$\sigma_{a_{11}}^2$ e $\sigma_{a_{22}}^2$: variâncias aditivas intrapopulacionais das populações 1 e 2, respectivamente.

$\sigma_{\tau_{12}}^2$: variância dos desvios dos efeitos aditivos intra (população 1) e interpopulacionais.

$Cov_{(a_1 \tau_{12})}$ e $Cov_{(a_2 \tau_{21})}$: covariâncias dos efeitos aditivos intrapopulacionais com os desvios dos efeitos aditivos intra e interpopulacionais.

Com base na expressão de $\sigma_{a_{12}}^2$, verifica-se que a SRR utiliza os efeitos aditivos intrapopulacionais (via $\sigma_{a_{11}}^2$) e os efeitos de dominância (via $\sigma_{\tau_{12}}^2$) nos locos em que as populações são divergentes e que apresentam dominância. Dessa forma, a SRR explora e aumenta a heterose dos cruzamentos. É importante notar que a SRR não utiliza a variância genética de dominância ($\sigma_{d_{(12)}}^2$).

Fazendo-se um paralelo entre a SRR e THS para o melhoramento do híbrido interpopulacional, tem-se, para efeito comparativo, que:

(i) SRR

Faz uso de h_{12}^2 e h_{21}^2 para as frações do ganho associados as recombinações nas populações 1 e 2, respectivamente.

(ii) THS

Faz uso de h_{12}^2 e $h_{21} h_{22} r_{G_{21}}$ para as frações do ganho associado à recombi-nação nas populações 1 e 2, respectivamente.

Verifica-se assim que, quando as herdabilidades intrapopulacional na po-pulação 2 (h_{22}^2) e interpopulacional (h_{21}^2) são iguais e a correlação $r_{G_{21}} = \sigma_{a_2 a_1} / (\sigma_{a_{22}} \sigma_{a_{21}})$ entre progêneres intrapopulacionais da população 2 e interpopulacio-nais for igual a 1, os esquemas SRR e THS conduzem ao mesmo ganho genéti-co. E isto ocorre quando as populações não são divergentes. Com divergência, a vantagem de THS sobre SRR deve advir então, da superioridade de h_{22}^2 sobre h_{21}^2 . A eficiência da THS sobre a SRR é diretamente proporcional a $h_{22} r_{G_{21}} / h_{21}$.

A escolha da população testadora não necessita, necessariamente, base-ar-se nas produtividades médias das populações. Em realidade, deve-se com-parar $h_{22} r_{G_{21}} / h_{21}$ com $h_{11} r_{G_{12}} / h_{12}$. Se $h_{22} r_{G_{21}} / h_{21} > h_{11} r_{G_{12}} / h_{12}$, a populaçao 2 deve ser utilizada como testadora. Caso contrário, a população 1 deve ser usa-da como testadora. Para haver eficiência do método THS sobre a SRR, uma das duas relações mencionadas ($h_{22} r_{G_{21}} / h_{21}$ ou $h_{11} r_{G_{12}} / h_{12}$) deve ser maior que 1.

Estimativas dos parâmetros herdabilidades intrapopulacionais (h_{11}^2 e h_{22}^2), herdabilidades interpopulacionais (h_{12}^2 e h_{21}^2) e correlações entre as performan-ces intrapopulacionais e interpopulacionais, também denominadas correlações entre capacidades gerais de combinação e capacidades gerais de hibridação ($r_{G_{12}}$ e $r_{G_{21}}$), são essenciais na escolha do melhor procedimento de seleção recorrente interpopulacional.

16.4. Índice de Seleção para a Seleção Recorrente Recíproca

A eficiência de um índice de seleção combinando as performances intra-populacional 2 e interpopulacional em relação ao uso de progêneres interpopula-cionais, para o melhoramento do híbrido interpopulacional, é dada por $\left[1 + \frac{(h_{22} r_{G_{21}} / h_{21} - r_{F_{21}})^2}{1 - r_{F_{21}}^2} \right]^{1/2}$, em que $r_{F_{21}}$ é a correlação fenotípica entre as perfor-mances intrapopulacional e interpopulacional. A eficiência deste índice em rela-ção ao uso apenas de progêneres intrapopulacionais da população 2 é dada por

$$[h_{21} / (h_{22} r_{G_{21}})] \left[1 + \frac{(h_{22} r_{G_{21}} / h_{21} - r_{F_{21}})^2}{1 - r_{F_{21}}^2} \right]^{1/2}.$$

Em espécies perenes, considerando a facilidade de obtenção de progêneres intrapopulacionais por polinização aberta, torna-se relevante verificar a eficiência do índice combinando as performances intrapopulacional e interpopulacional visando à seleção e à recombinação na população 1 (a mais produtiva).

Neste caso, a eficiência é dada por: $E = \left[1 + \frac{(h_{l1} r_{G12} / h_{l2} - r_{F12})^2}{1 - r_{F12}^2} \right]^{1/2}$.

Considerando que r_F refere-se à correlação fenotípica em plantas distintas tem-se que $r_{G12} = \frac{r_F}{h_{l1} h_{l2}}$ e, portanto, $r_F = r_{G12} h_{l1} h_{l2}$ (= herdabilidade do diferencial de seleção direto associado às progêneres interpopulacionais, via seleção indireta nas progêneres intrapopulacionais). Dessa forma:

$$E = \left[1 + \frac{(h_{l1} r_{G12} / h_{l2} - r_{G12} h_{l1} h_{l2})^2}{1 - (r_{G12} h_{l1} h_{l2})^2} \right]^{1/2}.$$

Considerando $h_{l1}^2 = h_{l2}^2 = 0,30$ e $r_{G12} = 0,80$ (o que denota que a seleção indireta é cerca de 20% menos eficiente), $E = 1,15$, ou seja, nesta situação, o uso simultâneo das informações das progêneres intrapopulacionais e interpopulacionais propicia 15% a mais de ganho genético na contribuição da população 1 para o melhoramento do híbrido interpopulacional.

O índice é dado por $I = b_1 X_{F_{12}} + b_2 X_{F_{11}}$, em que b_i refere-se aos pesos dos valores fenotípicos e $X_{F_{12}}$ e $X_{F_{11}}$ referem-se aos valores fenotípicos associados aos materiais interpopulacionais e intrapopulacionais, respectivamente. Atribuindo 1 a b_1 , b_2 é dado por:

$$\begin{aligned} b_2 &= \frac{\sigma_{F_{12}}}{\sigma_{F_{11}}} \cdot \frac{r_{G12} h_{l1} - r_{F_{12}} h_{l2}}{h_{l2} - r_{F_{12}} r_{G12} h_{l1}} \\ &= \frac{\sigma_{F_{12}}}{\sigma_{F_{11}}} \cdot \frac{r_{G12} h_{l1} - r_{G12} h_{l1} h_{l2}^2}{h_{l2} - r_{G12}^2 h_{l1}^2 h_{l2}}. \end{aligned}$$

Considerando $\sigma_{F_{12}} = \sigma_{F_{11}}$, tem-se $b_2 = 0,69$, para o presente exemplo. Neste caso, é dado à performance interpopulacional o peso 1 e peso 0,69, à performance intrapopulacional.

16.5. Seleção Recorrente em Algumas Espécies Perenes

São raras as publicações dedicadas exclusivamente ao melhoramento das várias espécies perenes, podendo-se citar apenas Ferwerda & Wit (1969), Cirad (1995) e Baudoim et al. (1997) que contemplaram o eucalipto, a seringueira, o cacau, o café-arábica, o café-robusta, o coco e o dendê.

A seguir são abordados aspectos relacionados com a seleção recorrente em algumas das principais espécies perenes relevantes cultivadas no Brasil. Os sistemas reprodutivos destas espécies são apresentados no capítulo inicial deste documento.

16.5.1. Espécies florestais

Eucalyptus

O eucalipto pertence à divisão *Angiospermae*, classe *Dicotyledonea*, ordem *Myrtales*, família *Myrtaceae* e gênero *Eucalyptus*. O gênero *Eucalyptus* possui cerca de 600 espécies, quase todas (exceto o *E. urophylla*, que ocorre na Indonésia, e o *E. deglupta*, que ocorre na Indonésia e em Papua Nova Guiné) nativas da Austrália. Dentre as espécies mais importantes para o Brasil, apenas duas apresentam ocorrência natural fora da Austrália: o *E. urophylla* que é nativo na Indonésia (principalmente nas Ilhas de Flores e de Timor) e o *E. tereticornis* que ocorre na Austrália e, também, em Papua Nova Guiné.

A Austrália foi descoberta em 1788 e é constituída por sete estados: Tasmania, Victoria, New South Wales, Queensland, Northern Territory, South Austrália e Western Austrália. A maioria das espécies economicamente importantes ocorrem nos quatro primeiros estados mencionados, embora o *E. camaldulensis* seja encontrado também nos outros três estados. A partir do descobrimento da Austrália, o Eucalyptus se espalhou rapidamente pelo mundo e, hoje, é plantado comercialmente em vários países, sendo que os maiores plantadores são: Índia (com cerca de 5 milhões de hectares), Brasil (com cerca de 3 milhões de hectares), África do Sul, Portugal, Espanha, China e Chile. As espécies mais plantadas mundialmente são o *E. urograndis*, *E. grandis*, *E. urophylla*, *E. globulus*, *E. camaldulensis* e *E. tereticornis*.

Uma classificação mais detalhada das principais espécies utilizadas no Brasil (e de outras não utilizadas no Brasil, mas empregadas pela indústria de celulose no mundo) é apresentada a seguir com base em Eldridge et al. (1993).

Subgênero *Sympyomyrthus*

Secção Transversaria : *E. grandis*, *E. saligna*, *E. urophylla*, *E. pellita*

Secção Maidenaria : *E. globulus*, *E. dunnii*, *E. nitens*, *E. smithii*

Secção Exsertaria : *E. camaldulensis*, *E. tereticornis*

Subgênero *Idiogenes* : *E. cloeziana*

Subgênero *Corymbia*: *E. citriodora*

Subgênero *Monocalyptus*: *E. pilularis*

As espécies *E. grandis*, *E. saligna*, *E. urophylla* e *E. pellita* são comumente empregadas pela indústria de celulose e cultivadas para este fim em regiões tropicais e subtropicais, assim como os seus híbridos. Destas, *E. grandis* e

E. saligna são espécies próximas, pertencentes a Série *Salignae* e *E.urophylla* e *E.pellita* são também espécies próximas. As espécies da Secção *Maidenaria* mencionadas são utilizadas pela indústria de celulose e cultivadas para este fim em regiões temperadas e subtropicais.

Para a produção de energia (carvão), as espécies mais recomendadas para regiões tropicais são *E. camaldulensis*, *E.urophylla*, *E. cloeziana* e *E. tereticornis*, as quais apresentam mais alta densidade da madeira. Destas, *E. camaldulensis* e *E. tereticornis* são espécies muito próximas. No Brasil, os materiais genéticos mais plantados para este fim são: clones híbridos de *E. urophylla* x *E. camaldulensis*, denominados de "E. urocum"; *E. urophylla*; *E. cloeziana*. Para a produção de óleo essencial uma das espécies mais plantadas é o *E. citriodora*.

As espécies geneticamente próximas *E. grandis* e *E. saligna*; *E. urophylla* e *E. pellita*; *E. camaldulensis* e *E. tereticornis* podem ser assim classificadas em termos de importância econômica e silvicultural:

- *E. grandis* (principal); *E. saligna* (secundária)
- *E. urophylla* (principal); *E. pellita* (secundária)
- *E. camaldulensis* (principal); *E. tereticornis* (secundária)

Em *Eucalyptus*, suspeita-se que os efeitos de dominância existam e apresentam baixa magnitude para os caracteres de crescimento (Hardner et al., 1996; Hodge et al., 1996; Borralho, 1997; Rezende & Resende 2000). Quanto à heterose propriamente dita, tem sido verificado que esta é de alta magnitude apenas em sítios marginais para ambas as espécies genitoras. Em sítios favoráveis a essas espécies, a superioridade do híbrido não tem sido confirmada (Martin, 1989; Eldridge et al., 1993). Assim, segundo Martin (1989), o vigor híbrido em *Eucalyptus* pode ser classificado como ambientalmente dependente.

Apesar de aparentemente baixa, a existência de dominância já é suficiente para propiciar uma leve vantagem da propagação clonal sobre a sexuada e para permitir a exploração da capacidade específica de combinação entre genitores, principalmente ao nível interespecífico. Dessa forma, híbridos de *Eucalyptus* têm sido utilizados em vários países, destacando-se as combinações *E. urophylla* x *E. grandis* no Brasil, África do Sul, Colômbia e Venezuela e *E. urophylla* x *E. grandis*, *E. grandis* x *E. tereticornis* e *E. urophylla* x *E. pellita* no Congo.

No âmbito mundial, a primeira estratégia de melhoramento de híbridos interespecíficos de *Eucalyptus* foi implementada a partir de 1989 no Congo (Vignerón, 1991; 1992), empregando-se a seleção recorrente recíproca baseada em cruzamentos fatoriais e cujos resultados iniciais foram publicados por Bouvet et al. (1992), Bouvet (1995) e Bouvet & Vignerón (1995). A estratégia de SRR empregada envolve os híbridos *E. urophylla* x *E. grandis* e *E. urophylla* x *E. pellita*. Para a população híbrida *E. urophylla* x *E. grandis*, a relação verificada para σ_d^2 / σ_a^2 do caráter volume aos 3 anos foi de 0,20 (Bouvet, 1995; Bouvet & Vignerón, 1995), concordando com o resultado de Rezende & Resende (2000).

Para o *E. urophylla* x *E. pellita*, a relação σ_d^2 / σ_a^2 mostrou-se mais elevada. Com base em análise crítica dos primeiros resultados obtidos no programa de melhoramento de híbridos de *Eucalyptus* no Congo, Bouvet (1995) concluiu pela adequabilidade da seleção recorrente recíproca baseada em famílias de irmãos germanos. Baudouim et al. (1997) relataram a obtenção de cinco cultivares híbridos que elevaram a produtividade média de 20 para 35 m³/ha/ano.

As considerações realizadas permitem concluir que a dominância em *Eucalyptus*, provavelmente, é suficiente para proporcionar efeitos consideráveis da capacidade específica de combinação envolvendo cruzamento entre genitores, concordando com Cotterill (1997) que descreve a ocorrência de efeitos de alta magnitude para a capacidade específica de combinação em *Eucalyptus globulus* (ao contrário do observado para *Pinus*, em que a capacidade geral de combinação dos genitores mostrou-se uma boa preditora da média da família de irmãos germanos obtidas do cruzamento entre eles), gerando famílias excepcionais de irmãos germanos, destinadas à propagação vegetativa.

No Brasil (na região tropical costeira do sul da Bahia e do norte do Espírito Santo, em áreas da Aracruz Celulose, Bahia Sul Celulose e Veracel Celulose), após intensa experimentação com várias espécies, procedências e clones de *Eucalyptus*, definiu-se como material genético ideal para atendimento do setor industrial, em termos quantitativos e qualitativos, clones do híbrido *E. urograndis*, ou seja, *E. urophylla* x *E. grandis*. Tal híbrido apresenta alto crescimento volumétrico na região e resistência ao cancro, além de gerar madeira de qualidade compatível com as exigências da fábrica. Este híbrido apresenta, também, boa taxa de enraizamento de estacas, as quais são utilizadas para o estabelecimento dos plantios comerciais clonais.

A geração contínua de clones superiores ao longo de vários ciclos seletivos ou gerações de melhoramento deve basear-se no melhoramento do híbrido interpopulacional de *E. urophylla* e *E. grandis*. As populações-base das espécies genitoras foram estabelecidas após intensivos testes de populações e progêniens de *E. grandis* e *E. urophylla*. Do *E. grandis*, foram descartadas populações tais quais as procedências da África do Sul, Zimbabwe e Coffs Harbour/Austrália, tendo-se selecionado indivíduos superiores de procedências da região de Atherton/Queensland/Austrália (750 m de altitude). Do *E. urophylla*, foram avaliadas populações das sete ilhas em que a espécie ocorre na Indonésia e selecionados indivíduos superiores, principalmente, das procedências Mount Lewotobi e Mount Egon da Ilha de Flores, além de procedências de baixa altitude das ilhas de Timor e Alor.

Pinus

As principais espécies de *Pinus* de interesse para produção de celulose de fibra longa e madeira serrada no Brasil são originárias da América Central, do México e dos Estados Unidos. Dentre as espécies mais plantadas destacam-se: (i) subtropicais: *Pinus taeda* (espécie de *Pinus* mais plantada no Brasil) e *Pinus*

elliotii (utilizada também para a extração de resina); (ii) tropicais: *Pinus caribaea* var. *hondurensis* (representa cerca de 70% dos plantios com pinus tropicais) e *Pinus oocarpa* (representa cerca de 15% dos plantios com pinus tropicais).

Em *Pinus*, os efeitos de dominância são, aparentemente, de baixa magnitude para características de crescimento (volume de madeira, altura e diâmetro das árvores), tendo-se constatado heterose praticamente nula (Cotterill et al., 1987; Cotterill, 1997). Segundo Cotterill (1997), os efeitos da capacidade específica de combinação não são importantes em *Pinus* e a capacidade geral de combinação dos genitores é uma boa preditora da média do cruzamento entre eles.

Com base no exposto e considerando que os plantios clonais (que permitem a fixação imediata de genótipos heteróticos) não são ainda muito praticados em *Pinus*, as estratégias de melhoramento são, em geral, baseadas em SRI (conforme revisão apresentada em Resende, 1999a).

Acácia-negra (*Acacia mearnsii*)

A acácia-negra é nativa da Austrália e é cultivada, principalmente, na África do Sul, no Brasil e na China. É utilizada para extração de tanino (empregado no curtimento de couros e peles) e sua madeira é utilizada pela indústria de celulose, dentre outros usos. No Brasil é plantada no Rio Grande do Sul e os principais caracteres de interesse para o melhoramento são a produção de madeira e de tanino, e a resistência à doença gomose e ao inseto serrador (*Oncideres impluviata*). Dentre as populações mais produtivas no Brasil destaca-se a de Bateman's Bay.

Para a acácia-negra não existem ainda informações sobre os níveis de dominância associados aos principais (produção de tanino e de madeira) caracteres de importância econômica. Este fato, associado à não utilização de plantios clonais operacionais, tem conduzido a adoção de programas de SRI (Higa & Resende, 1992), no Brasil, na África do Sul e na China.

Acacia mangium* e *Acacia auriculiformis

Acacia mangium, *Acacia auriculiformis* e *Acacia mearnsii* são as três espécies (de origem Australiana) do gênero *Acacia*, que apresentam madeira de comprovada qualidade (comparável à de *Eucalyptus*) para celulose de fibra curta e são plantadas comercialmente para este fim, principalmente na Ásia. As espécies *Acacia mangium* e *Acacia auriculiformis* e seus híbridos são cultivados, principalmente, no Sudeste Asiático, nos países Malásia, Tailândia, Vietnã, Indonésia e Filipinas. Híbridos de *A. mangium* x *A. auriculiformis* têm apresentando heterose para crescimento e produtividade de polpa celulósica e valores intermediários das duas espécies, para densidade da madeira (Kha et al., 1998). Intensos programas de melhoramento intrapopulacional têm sido realizados em ambas as espécies, tanto com vistas ao melhoramento destas quanto do híbrido interespecífico (Hardiyanto, 1998).

Seringueira (*Hevea brasiliensis*)

A seringueira é uma das principais espécies florestais nativas do Brasil. Sua ocorrência natural está restrita à região amazônica, no Brasil, na Bolívia, no Peru, no Equador, na Colômbia, na Venezuela, no Suriname e na Guiana. A maior área de cultivo da seringueira localiza-se no Sudeste Asiático, região que detém 90% da produção mundial. A contribuição do Brasil está abaixo de 1%. Os principais países produtores são pela ordem: Tailândia, Indonésia, Malásia, Índia, China, Filipinas e Vietnã. No Brasil, as principais áreas produtoras denominam-se áreas de escape à doença denominada mal-das-folhas (*Microcyclus ulci*) e localizam-se nos Estados de São Paulo (maior produtor brasileiro), Mato Grosso, Mato Grosso do Sul, sul da Bahia e noroeste do Paraná.

Nos programas de melhoramento da seringueira são enfatizados os caracteres produção de borracha, vigor avaliado por meio da circunferência do caule, resistência ao vento, morfologia da casca (espessura e grau de rugosidade), número de anéis de vasos laticíferos e resistência à doenças. Simmonds (1989) relata que os caracteres produção de borracha e precocidade de produção são os mais importantes e que uma alta produção de borracha é conseguida por meio de alto vigor da árvore, alta participação de assimilados em borracha, retenção do crescimento e vigor sob sangria, bem como resistência ao vento e a doenças. A precocidade de produção depende do vigor da planta na fase jovem. Segundo esse autor, os caracteres qualitativos da borracha são menos importantes em termos de melhoramento genético, visto que podem ser alterados via tecnologia pós-colheita.

Programas de melhoramento genético da seringueira têm sido realizados, principalmente no Brasil, na Malásia e na Costa do Marfim, e relatos a esse respeito podem ser encontrados em Gonçalves et al. (1983), Paiva & Kalil Filho (1985), Gonçalves (1986), Paiva & Gonçalves (1989), Simmonds (1989), Gonçalves et al. (1995), Clement-Demange et al. (1995), Resende et al. (1996d), Resende (1999a) e Costa et al. (2000).

Os programas de melhoramento genético da seringueira no Brasil enfatizam, principalmente, os caracteres produção de borracha, resistência ao mal-das-folhas e vigor (circunferência do caule). Para o caráter produção de borracha, Resende et al. (1996d) apresentaram os valores médios de literatura de 0,20 e 0,30 para os parâmetros herdabilidades individuais no sentido restrito e amplo, respectivamente. Estes valores revelam a existência de dominância e provável heterose para o caráter, concordando com os relatos de Paiva & Gonçalves (1989).

Os plantios clonais são a regra em seringueira. Entretanto, programas de seleção recorrente recíproca não foram ainda implementados nessa espécie (Baudouim et al., 1997), provavelmente devido à utilização, por longo tempo (desde 1876), de uma única população de base genética estreita nos programas de melhoramento de todo o mundo. Recentemente novas introduções (materiais silvestres) foram realizadas, mas o nível de produção dessas populações é ainda muito baixo (cerca de 10% daquele observado para a população melhorada),

fato que tem demandado a adoção prévia de programas de SRI nestas populações para, posteriormente, utilizá-las em cruzamento (Baudouim et al., 1997).

Leucena (*Leucaena leucocephala*)

A leucena é uma das mais importantes espécies do grupo das leguminosas arbóreas, apresentando as seguintes características favoráveis: rápido crescimento; usos múltiplos e facilidade para consórcio com culturas agrícolas; fixadora de nitrogênio e melhoradora do solo; sistema radicular robusto, tolerância à seca e à salinidade; poucos problemas relacionados a pragas e doenças. É utilizada como madeira, como melhoradora do solo e na alimentação animal.

O gênero *Leucaena* é originário principalmente da América do Sul e da América Central e concentra várias espécies, destacando-se a *Leucaena leucocephala* que apresenta maior produção de madeira. Essa espécie é autógama e apresenta três tipos quanto ao hábito de crescimento: havaiano (hábito arbustivo); peruano (tipo arbóreo, grande produtor de forragem); salvadorenho (tipo arbóreo gigante, grande produtor de madeira).

Os caracteres de interesse do melhoramento em leucena são: produção de forragem, produção de madeira, tolerância à acidez do solo e ao frio. Os programas de melhoramento, em geral, destacam a hibridação entre linhagens e também entre espécies, mas não há relato do uso prático de programas de seleção recorrente. Estratégias de melhoramento para *Leucaena leucocephala* são discutidas por Resende & Medrado (1994).

16.5.2. Espécies produtoras de alimentos estimulantes

Erva-mate (*Ilex paraguariensis*)

A erva-mate é nativa da Argentina, do Paraguai e do Brasil, nos Estados do Rio Grande do Sul, Santa Catarina, Paraná e Mato Grosso do Sul. Esta espécie é cultivada nestes mesmos locais como fonte de bebidas do tipo chá e chimarrão. Os programas de melhoramento da erva-mate concentram-se nos seguintes caracteres: produção de massa foliar, qualidade da bebida, resistência a insetos praga, arquitetura de copa e persistência foliar.

Para a erva-mate não existem ainda informações sobre o nível de dominância do caráter produção de massa verde e sobre a adequabilidade prática da propagação vegetativa. Estes fatores, associados à recente domesticação da espécie, têm levado à adoção de programas de SRI (Resende et al., 1995a, 1997a, 2000b; Belingheri & Kricun, 1997).

Chá-da-Índia (*Camellia sinensis*)

O chá-da-Índia é originário da China e fornece a bebida estimulante mais consumida no mundo. Por volta de 1700, variedades de chá foram introduzidas

na Indonésia e mais tarde na África. No século XIX, plantações em larga escala foram estabelecidas na Índia. Na América do Sul existem plantações principalmente na Argentina (em regiões onde também se cultiva a erva-mate) e no Vale do Ribeira-SP, Brasil.

Os programas de melhoramento genético do chá iniciaram-se por volta de 1910 na Indonésia, na África, na China e na Índia, portanto, bem antes do desenvolvimento da Genética Quantitativa e da Estatística Experimental. Em função disto, as técnicas empregadas (seleção seguida de clonagem) conduziram a um grande estreitamento da base genética à semelhança do que ocorreu com a seringueira. Dessa forma, recentemente, os programas têm priorizado a introdução de novos germoplasmas e a prática da SRI (Resende et al., 1998b).

Cacau (*Theobroma cacao*)

O cacau é originário da América do Sul (região Amazônica) e América Central e existem quatro principais tipos de população: Forastero do Alto Amazonas, Forastero do Baixo Amazonas, Criollo e Trinitário (cruzamento entre os tipos Criollo e Forastero). Os maiores plantadores mundiais de cacau são: Costa do Marfim, Gana, Brasil, Nigéria, Camarões, Malásia e Indonésia. Importantes programas de melhoramento do cacau são realizados na Costa do Marfim, no Brasil, em Trinidad, na Malásia, em Gana, Camarões e Nigéria.

As variedades cultivadas de cacau são geralmente propagadas por meio de sementes híbridas (Dias, 1998a; Pereira et al., 1999), a fim de capitalizar a heterose, a qual tem sido observada para os principais caracteres: produção, componentes de produção, vigor, resistência a doença e precocidade (Dias, 1998a). No programa de melhoramento realizado na Costa do Marfim, as quatro populações (tipo) foram divididas em dois grupos (Forastero Baixo Amazonas junto com Trinitário, Forastero Alto Amazonas junto com Criollo), sendo que dois ciclos de SRI serão desenvolvidos dentro de cada grupo e, posteriormente, será iniciado um programa de SRR com os dois grupos (Clement-Demange et al., 1993; Baudouim et al., 1997). O uso dos tipos Criollo e Trinitário em cruzamento tem como objetivo a introdução de genes para qualidade do cacau no germeplasma Forastero. Descrições detalhadas sobre estratégias e métodos de seleção no melhoramento genético do cacau são apresentadas por Paulin (1995) e Dias e Resende (2001b).

Café-arábica (*Coffea arabica*)

O café é uma das principais fontes de divisas para o Brasil. As duas principais espécies de cafeiro cultivadas no mundo são o café-arábica (*Coffea arabica*) e o café-robusta (*Coffea canephora*), as quais respondem por cerca de 70% e 30% da produção mundial. No Brasil, a principal espécie plantada é o café-arábica (sobretudo nos Estados de Minas Gerais, São Paulo e Paraná).

Coffea arabica apresenta distribuição geográfica natural restrita aos países Etiópia (região sudeste), Sudão (região sudeste) e Quênia (região norte), em altitudes entre 1.000 e 2.000 metros. O café arábica é cultivado, principalmente, no Brasil (maior produtor mundial), Colômbia, Quênia e Índia. Os principais caracteres de interesse ao melhoramento do cafeiro são: produção de frutos, qualidade da bebida, resistência a pragas e doenças, diâmetro do caule e número de ramos plagiotrópicos. Estes dois últimos caracteres têm-se mostrado correlacionados geneticamente com a produção de café e seus componentes.

O café arábica é uma espécie autógama, fato que tem orientado a adoção dos métodos de seleção de linhas em plantios comerciais e método genealógico (Mendes, 1999). Estudos recentes têm demonstrado a significância dos efeitos da capacidade específica de combinação e a importância da utilização de híbridos (Cilas et al., 1998a). No Brasil, a seleção recorrente intrapopulacional foi proposta como alternativa viável para o melhoramento desta espécie (Ramalho et al., 1999).

Café-robusta (*Coffea canephora*)

Coffea canephora tem ampla distribuição geográfica, ocorrendo em regiões de pequena altitude no oeste e centro do continente africano. O café-robusta é cultivado principalmente na Indonésia, no Vietnã e na Costa do Marfim. No Brasil, essa espécie apresenta área de plantio mais restrita (plantado principalmente nos Estados do Espírito Santo, Rondônia e sul da Bahia). Os maiores produtores mundiais de café-robusta são Indonésia, Vietnã e Brasil.

As principais variedades de café-robusta utilizadas na África são híbridos advindos do cruzamento entre dois grupos genéticos (variedades da Guiné ou da África Ocidental, e variedades Congolesas ou da África Central), com características complementares. Em função disto, um programa de seleção recorrente recíproca vem sendo desenvolvido, desde 1984, na Costa do Marfim, o qual conduziu a um ganho genético de 60% para o caráter produção de café. Este resultado confirma a adequabilidade da SRR (Baudouim et al., 1997). No Brasil, a seleção recorrente intrapopulacional e a seleção de clones visando à propagação vegetativa têm sido realizadas com sucesso no Estado do Espírito Santo (Ferrão et al., 1999).

16.5.3. Palmáceas

Coco (*Cocos nucifera*)

O Sudeste Asiático é, provavelmente, a região de origem do coqueiro, de onde foi introduzido na Índia. Posteriormente, a espécie foi levada para o continente africano e, deste, para as Américas. Os principais países produtores são: Indonésia, Filipinas e Índia.

O coqueiro possui mais de 300 ecótipos e duas variedades principais: o coqueiro gigante (variedade alógama) e o coqueiro-anão (variedade autógama).

Este último apresenta as subvariedades verde, vermelha e amarela (recessiva em relação à verde e à vermelha). A subvariedade vermelha apresenta dois tipos distintos fenotipicamente: o anão-da-malásia e o anão-dos-camarões. Do ponto de vista do melhoramento genético, os caracteres mais importantes são o número de frutos, a quantidade de copra no fruto, a resistência a pragas e doenças, e a tolerância ao déficit hídrico (Siqueira et al. 1998).

Embora a clonagem para plantios comerciais em coqueiro não seja ainda uma tecnologia disponível, a utilização de sementes híbridas para plantios em larga escala é corriqueira. Isto se deve à facilidade de obtenção dessas sementes por meio do uso das variedades anãs (autógamas) como genitores femininos e das técnicas de emasculação, conservação de pólen e polinização controlada.

Devido à superioridade dos cultivares híbridos, o programa de melhoramento do coco na África baseia-se em um esquema de SRR que visa melhorar simultaneamente os híbridos Anão x Gigante e Gigante x Gigante, envolvendo dois ecótipos do coqueiro gigante (Baudouim et al., 1997).

Dendê (*Elaeis guineensis*)

O dendê é originário da África e tem sido cultivado no sul da Bahia e no norte do Brasil, visando à produção de óleo. A produção do dendê concentra-se, principalmente, nos países do Sudeste Asiático, sendo a Malásia e a Indonésia os maiores produtores mundiais. Outros países que cultivam esta espécie são a Costa do Marfim e Camarões.

Os principais programas de melhoramento genético do dendê são realizados na Costa do Marfim, na Malásia e no Brasil. O programa realizado pelo Brasil baseia-se em germoplasma introduzido pelo Cirad (França) e visa, dentre outras características, ao aumento da produção de óleo (principalmente por meio do aumento do número e peso de cachos de frutos), à redução do porte, à resistência ao amarelecimento fatal (o germoplasma do caiauá - *Elaeis oleifera* - é também utilizado como fonte de resistência) (Barcelos et al. 2000).

As variedades comerciais de dendê são produzidas por polinizações de duas populações distintas: uma utilizada como genitor feminino (*Deli dura*) e a outra como genitor masculino (*African pisifera*), a fim de conciliar características agronômicas favoráveis. Dessa forma, um programa de SRR foi adotado visando ao melhoramento do dendê na África (Bourdeix et al., 1991 a e b; Baudouim et al., 1997). Em realidade, o programa de seleção recorrente em dendê foi iniciado por volta de 1950 (Meunier & Gascon, 1972) e tem produzido excelentes resultados (Cochard et al., 1993; Baudouim, 1995). Um tratamento histórico e prático do melhoramento genético do dendê foi realizado por Hartley (1977).

Pupunha (*Bactris gasipaes*)

A pupunheira é uma palmeira nativa da Amazônia e da América Central. Existem populações com e sem espinhos, das quais as populações inermes são

preferidas para cultivo. Populações sem espinhos ocorrem no Peru, no Brasil, na Colômbia e na Costa Rica. As principais populações sem espinhos cultivadas são a Yurimágua (Peru), San Carlo (Costa Rica) e Benjamim Constant (Brasil). A pupunha é, também, classificada em raças quanto ao tamanho do fruto: macrocarpa, mesocarpa e microcarpa.

O maior uso da pupunha por habitantes da Amazônia é o fruto, consumido diretamente após o cozimento em água com sal, ou no preparo de várias iguarias. Os frutos são também utilizados para ração animal e para a obtenção de farinha que é empregada na fabricação de pães e bolos. Porém, atualmente, o palmito é o principal motivo do plantio da pupunha em larga escala. Os principais interesses pelo seu cultivo são, além de ser uma cultura perene: crescimento a pleno sol, precocidade, rusticidade, perfilhamento, palatabilidade e não escurecimento do palmito após o corte.

Os mais importantes caracteres de interesse do melhoramento são: número de perfilhos, diâmetro da planta à altura do colo, altura da folha bandeira, produção líquida de palmito, tolerância ao déficit hídrico e à geada (em algumas regiões de plantio). Variabilidade genética para os caracteres de interesse tem sido detectada, dependendo da população utilizada (Clement & Bovi, 1999; Farias Neto & Resende, 2001).

Quanto aos processos de seleção empregados no melhoramento da pupunheira, destaca-se a SRI. Entretanto, Clement et al. (1987) argumentam em favor do uso da SRR, a exemplo do que vem sendo amplamente realizado pelos franceses com o dendê. Os referidos autores justificam a proposição com base em uma provável divergência genética entre as três principais populações sem espinhos utilizadas, as quais são originárias de diferentes países.

16.5.4. Fruteiras

Caju (*Anacardium occidentale*)

O cajueiro é uma fruteira nativa do Brasil. É cultivada em vários países tropicais como fonte de castanha e suco. Os principais países produtores são: Índia, Brasil, Moçambique, Tanzânia e Quênia. Países emergentes no plantio do caju são: Costa do Marfim, Benin, Guiné-Bissau, Indonésia, Tailândia e Vietnã.

Os caracteres importantes no contexto do melhoramento do cajueiro são o número de frutos por planta, a produtividade de castanhas (número e peso), o porte da planta, o peso da amêndoia e a resistência a doenças. Existem dois tipos principais de cajueiros cultivados: comum e anão-precoce (Barros e Crisóstomo, 1995).

Os plantios comerciais do cajueiro são realizados via clonagem. Este fato, associado à existência de heterose para os caracteres produtivos (Damodaran, 1975; Manoj & George, 1993; Barros et al., 1999; Crisóstomo et al., 1999; Cavalcanti et al., 2000), tem estimulado a aplicação de esquemas de SRR nesta espécie, em populações dos tipos anão e comum.

Acerola (*Malpighia emarginata*)

A aceroleira é originária das Antilhas e foi introduzida no Brasil na década de 1950. É cultivada comercialmente em Porto Rico, no Hawaí, em Cuba, na Flórida (EUA) e no Brasil. É cultivada em vários estados brasileiros, desde o Paraná até Rondônia. O fruto da aceroleira desperta grande interesse comercial, pelo elevado teor de vitamina C que apresenta.

Os principais caracteres da aceroleira, do ponto de vista do melhoramento, são: produtividade de frutos, arquitetura da copa, teor de vitamina C, tamanho do fruto, resistência a doenças. Os programas de melhoramento desta espécie são recentes no Brasil e concentram-se na seleção recorrente intrapopulacional (Gonzaga Neto, 1995, 1999; Oliveira & Soares Filho, 1995, 1999; Paiva et al., 1999; Paiva et al., 2001b).

Maçã (*Malus domestica*)

Dentre as fruteiras de clima temperado, a macieira ocupa a segunda posição em importância comercial no Brasil, superada apenas pela videira. A macieira é originária da Europa e da Ásia. No Brasil, os principais caracteres para o melhoramento são a produtividade de frutos, a baixa exigência em frio hibernal, a precocidade de frutificação e a resistência a pragas e doenças (Denardi & Camilo, 1998).

Para a maioria dos caracteres agronômicos da maçã, os efeitos aditivos parecem ser muito mais importantes do que os não aditivos. Assim, programas de SRI poderiam ser especialmente indicados (Noiton & Shelbourne, 1992; Durel et al., 1998).

Cupuaçu (*Theobroma grandiflorum*)

O cupuaçuzeiro é uma fruteira nativa da Amazônia, pertencente ao mesmo gênero do cacauzeiro. Nesta espécie, os principais caracteres para o melhoramento são: número de frutos por planta, peso de frutos, porcentagem de polpa e resistência à vassoura-de-bruxa. Programas de seleção recorrente intrapopulacional para o melhoramento destes caracteres estão sendo realizados no Brasil (Souza et al., 1998; Souza & Silva, 2000).

Com base nos programas de melhoramento de espécies perenes citados, pode-se inferir que a SRR é recomendada sobretudo quando os propágulos melhorados a serem utilizados comercialmente são híbridos, e/ou clones, e/ou quando os efeitos da capacidade específica de combinação são significativos. Dessa forma, espécies como *Eucalyptus*, seringueira, café-canefora, cacau, caju, coco e dendê são particularmente favoráveis a programas de SRR.

16.6. Esquema Detalhado para o Melhoramento de Híbridos Interpopulacionais ou Interespecíficos

A seguir será descrito o esquema de seleção recorrente recíproca (com famílias de meios-irmãos) para espécies perenes, proposto por Resende & Higa (1990), conforme apresentado na Fig. 16.1.

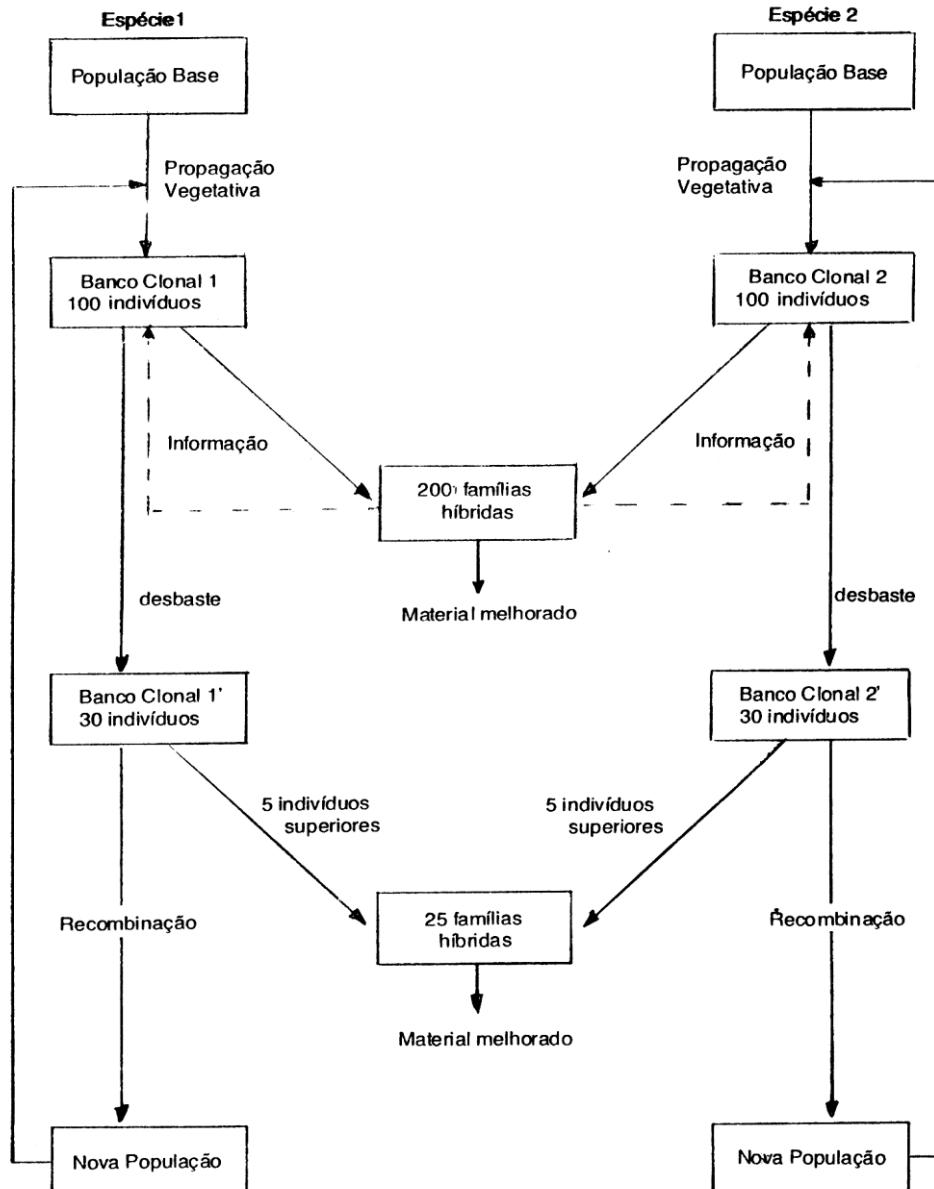


Fig. 16.1. Esquema de melhoramento proposto para a seleção de híbridos interespecíficos ou interpopulacionais

As seguintes etapas estão envolvidas em cada ciclo de seleção:

Etapa 1. Formação da população inicial de cada espécie.

A população inicial de cada espécie deverá ser constituída por 100 indivíduos superiores (selecionados com base em valores genéticos preditos) obtidos a partir de populações com maior grau de melhoramento. Esses indivíduos deverão ser transferidos para um banco clonal por enxertia.

Etapa 2 - Obtenção de híbridos interespecíficos para avaliação da capacidade de combinação dos genitores.

Os 100 indivíduos de cada espécie deverão ser utilizados como genitores femininos em cruzamentos com mistura de pólen da espécie recíproca, gerando 200 famílias de híbridos interespecíficos.

Etapa 3 - Avaliação das famílias híbridas e seleção de genitores.

As 200 famílias obtidas deverão ser avaliadas em ensaios com repetições em vários locais. Com base nos resultados dos ensaios, deverão ser selecionados os 30 genitores de cada espécie com maior capacidade de combinação com a espécie recíproca, desbastando-se os 70 indivíduos não selecionados no banco clonal. Os indivíduos superiores dentro dos ensaios deverão ser clonados, avaliados em testes clonais e os melhores propagados, vegetativamente, para plantios comerciais.

Etapa 4 - Recombinação do 1º ciclo de seleção e obtenção de famílias híbridas para avaliação, visando explorar a capacidade específica de combinação.

Os 5 indivíduos com maior capacidade geral de hibridação (os quais produziram os melhores híbridos) de uma espécie deverão ser cruzados com os 5 superiores da espécie recíproca, em todas as combinações possíveis, produzindo 25 famílias de irmãos germanos híbridos. Este procedimento permitirá selecionar famílias com performances excepcionalmente boas, com aproveitamento máximo da heterose através da capacidade específica de combinação. Assim, estas famílias deverão ser avaliadas em ensaios com repetições em vários locais e as superiores deverão ser propagadas para plantio comercial. Novamente, os indivíduos superiores dentro dos ensaios deverão ser clonados, avaliados em testes clonais e os melhores propagados vegetativamente para plantios comerciais.

No banco clonal correspondente a cada espécie, deverão ser colhidas sementes de polinização aberta entre os 30 indivíduos selecionados, para recombinação do 1º ciclo de seleção.

Etapa 5 - Plantio das populações recombinadas de cada espécie e seleção de indivíduos superiores para clonagem.

A partir das sementes de polinização aberta, colhidas nos bancos clonais correspondentes a cada espécie, serão geradas populações experimentais dentro das quais deverão ser selecionados 100 indivíduos superiores a serem clonados para início do 2º ciclo de seleção, ou seja, volta-se à Etapa 1. Na seleção desses indivíduos, poderá ser considerado o pedigree, como forma de evitar o rápido incremento da endogamia na população.

Para melhor compreensão do método, a Fig. 16.1 ilustra de maneira sucinta as etapas envolvidas. Esquema semelhante (exceto pela exploração da capacidade específica de combinação das 25 famílias de irmãos germanos) foi apresentado por Baradat et al. (1995), com os mesmos números de indivíduos-base (100) e a serem selecionados e recombinados (30), com base na capacidade geral de combinação. O esquema apresentado por Baradat et al. (1995) também difere em termos do delineamento utilizado (cruzamentos fatoriais).

A utilização das famílias de irmãos germanos superiores para plantio comercial, conforme descrito na Etapa 4, poderá ser realizada opcionalmente de duas maneiras: via sementes ou propagação vegetativa. Tais famílias superiores somente poderão ser propagadas via sementes se os genitores que as originaram forem pronunciadamente auto-estéreis, apresentando, em delineamento de pomar, baixa taxa de autofecundação. A propagação via sementes seria interessante devido à grande quantidade de material genético melhorado produzido a baixo custo. Assim, a seleção de aproximadamente três famílias superiores aumentará as chances de que pelo menos uma delas possa ser propagada via sementes. Entretanto, na impossibilidade de utilizar a propagação via sementes sugere-se que as famílias de irmãos germanos superiores sejam propagadas vegetativamente em massa, ou seja, em escala logarítmica, conforme apresentado por Cotterill et al. (1987).

A utilização de famílias, em vez de clones, é desejável quando se objetiva a obtenção de materiais mais estáveis fenotipicamente através dos ambientes e menos vulneráveis a fatores adversos (condições climáticas estressantes e ocorrência de pragas e doenças). As famílias a serem utilizadas comercialmente poderão apresentar uma combinação adequada e balanceada de diferentes genótipos que lhes conferirão boas performances e maior estabilidade, em função da homeostase populacional conferida pela interação compensatória de indivíduos dentro dessas famílias. Por outro lado, o emprego de famílias de irmãos germanos conduz a plantios mais homogêneos do que aqueles obtidos a partir de sementes de polinização aberta dos melhores genitores, já que, dentro de famílias de irmãos germanos, há menor variabilidade genética que dentro de famílias de meios-irmãos.

Em cada ciclo, as famílias de irmãos germanos geradas (etapa 4) deverão apresentar elevadas médias em relação à média da população híbrida (200 famílias) resultante do cruzamento entre as duas populações, pois a sua obtenção terá sido orientada com base na capacidade geral de hibridação dos genitores. A seleção entre famílias de irmãos germanos permite um ganho genético dado por:

$G_s = d_s [(1/2) \hat{\sigma}_{a(12)}^2 + (1/4) \hat{\sigma}_{d(12)}^2] / \hat{\sigma}_{F(12)}^2$, em que:

d_s : diferencial de seleção.

$\hat{\sigma}_{a(12)}^2$: variância genética aditiva interpopulacional.

$\hat{\sigma}_{d(12)}^2$: variância genética de dominância interpopulacional.

$\hat{\sigma}_{F(12)}^2$: variância fenotípica entre médias de famílias de irmãos germanos.

Este progresso genético será alto devido à considerável variabilidade genética explorada pelo método e à alta pressão de seleção (1-3 famílias selecionadas), a qual, provavelmente, levará a um alto diferencial de seleção.

De um ciclo para outro, os 200 híbridos gerados (Etapa 2) capitalizarão um ganho genético dado por (conforme esquema de setas apresentado na Fig. 16.2).

$G_s = k_1 (1/4) \sigma_{a(12)}^2 / \sigma_{F(12)}^2 + k_2 (1/4) \sigma_{a(21)}^2 / \sigma_{F(21)}^2$ em relação aos 200 híbridos do ciclo anterior, onde:

$\sigma_{a(12)}^2$ e $\sigma_{a(21)}^2$: variâncias genéticas aditivas interpopulacionais que surgem quando as espécies 2 e 1 são utilizadas como doadoras de pólen nos cruzamentos, respectivamente.

$\sigma_{F(12)}$ e $\sigma_{F(21)}$: desvios padrões fenotípicos entre famílias de meios-irmãos, quando as espécies 2 e 1 são utilizadas como doadoras de pólen nos cruzamentos, respectivamente.

k_1 e k_2 : diferenciais de seleção padronizados referentes às seleções de genitores nas espécies 1 e 2, respectivamente.

Assim, híbridos superiores serão gerados a cada ciclo de seleção recorrente recíproca. Além da possibilidade de utilização das famílias híbridas superiores, a seleção recorrente recíproca possibilita a clonagem de indivíduos híbridos superiores a cada ciclo, justificando a sua utilização.

Em cada ciclo de seleção, a magnitude da variância genética aditiva interspecífica deverá ser avaliada e comparada com as dos ciclos anteriores. Quando for constatada uma grande redução de variabilidade, deverão ser injetados novos indivíduos superiores provenientes de outras populações divergentes da mesma espécie.

Os híbridos superiores gerados com o processo de seleção poderão ser reunidos na forma de um composto de maneira a produzir populações com médias provavelmente altas e ampla variabilidade genética, portanto, promissoras para novos programas de seleção.

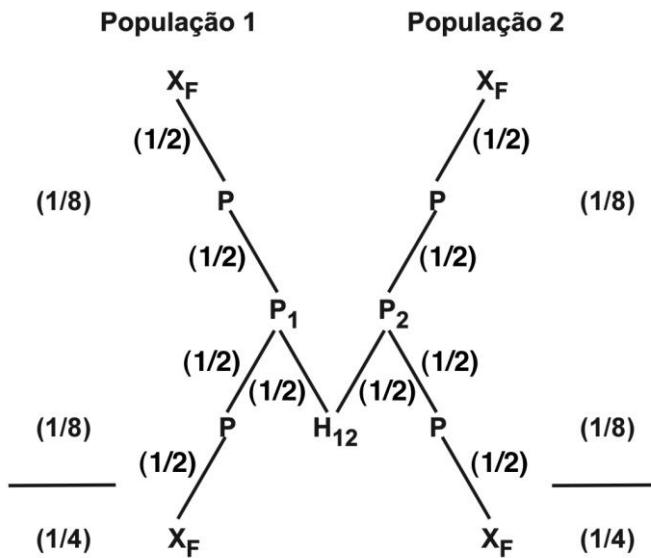


Fig. 16.2. Esquema representativo das covariâncias genéticas aditivas no numerador do ganho com a seleção recíproca com famílias de meios-irmãos, onde X_F , P , P_i e H_{12} referem-se às médias das progênies híbridas, genitores, descendentes dos genitores selecionados após a recombinação e híbrido melhorado, respectivamente.

16.7. Melhoramento de Híbridos Triplos

Híbridos triplos são ocasionalmente usados em espécies perenes (Nikles, 1992). Considerando o híbrido triplo $(A \times B) \times C$, tem-se a seguinte composição genética da média do híbrido:

$$\begin{aligned}\bar{H} &= (1/4) A + (1/4) B + (1/2) C + H_e \\ &= (1/4) g_A + (1/4) g_B + (1/2) g_c + (1/2) (s_{AC} + s_{BC})\end{aligned}\quad \text{em que:}$$

H_e : heterose.

g e s : capacidade geral e específica (heterose) de combinação.

Com base em \bar{H} , tem-se que o melhor preditor de \bar{H} é:

$$\bar{H} = (1/2)(A \times C) + (1/2)(B \times C)$$

Dessa forma, o melhor esquema para melhoramento do híbrido triplo é a SRR conduzida nos pares de população $A \times C$ e $B \times C$, devendo-se recombinar os genitores em A com base na seleção baseada em $A \times C$, recombinar os genitores em B com base na seleção baseada em $B \times C$ e recombinar os genitores em C com base na média de $A \times C$ e $B \times C$. Tal híbrido quando produzido cruzando diretamente $(A \times C) \times (B \times C)$, perde 50% da heterose (em relação aos genitores $(A \times B)$ e C) verificada em $(A \times B) \times C$, pois $(A \times C) \times (B \times C)$ é o F_2 do híbrido triplo. Assim, para a produção do triplo deve-se cruzar os selecionados em A e B produzindo $(A \times B)$ e então cruzar $(A \times B)$ com os selecionados em C .

Pela expressão de \bar{H} , em que a média (\bar{H}) do híbrido triplo é dada pela média dos dois híbridos simples, pode-se inferir que o híbrido triplo não capitaliza a heterose entre as espécies A e B (heterose do híbrido A x B). Assim, se o objetivo for apenas a obtenção de heterose para um caráter, não se justifica a utilização dos híbridos triplos, bastando trabalhar com o melhor híbrido simples dentre A x C e B x C.

Dessa forma, pode-se dizer que o uso do híbrido triplo só é justificado quando se objetiva complementariedade, ou seja, quando se deseja reunir características distintas de três espécies em um só indivíduo. Por outro lado, se existe uma população (A x B) já estabilizada (em equilíbrio, após pelo menos duas recombinações) e esta apresenta maior média que ambas, A e B, o uso de (A x B) x C seria plenamente justificado.

É importante frisar que o desempenho do híbrido triplo pode não depender de apenas um caráter. Considere, por exemplo, o híbrido triplo *E. urophylla* x (*E. camaldulensis* x *E. grandis*). Provavelmente o híbrido *E. urophylla* x *E. grandis* apresente maior crescimento que *E. urophylla* x *E. camaldulensis*, pois as duas primeiras devem ter melhores alelos para crescimento do que o *E. camaldulensis*. Entretanto, o *E. camaldulensis* tem maior resistência à seca e seus alelos podem ser essenciais para determinados ambientes de estresse. Assim, a produtividade final dependerá de uma série de caracteres desejáveis que depende de alelos das três espécies simultaneamente (efeito da complementariedade na heterose) para o caráter produtividade. Dessa forma, somente resultados experimentais poderão indicar, com segurança, a melhor estratégia a ser adotada.

16.8. Híbridos Interespecíficos Importantes no Gênero *Eucalyptus*

A maioria dos materiais genéticos de *Eucalyptus* utilizados comercialmente nos plantios destinados à produção de celulose e papel são clones híbridos, os quais apresentam heterose para crescimento e níveis adequados quanto a vários caracteres componentes da qualidade da madeira, do rendimento de celulose e do baixo custo de produção de celulose. Dessa forma, os programas de hibridação e de melhoramento interespecífico de híbridos são essenciais às empresas de produção de celulose de fibra curta, que se baseiam na eucaliptocultura.

Dentre as espécies puras de *Eucalyptus* mais aptas à produção econômica de celulose destacam-se o *Eucalyptus grandis* e o *E. globulus*. Outras que apresentam bons resultados quando em hibridação são: *E. urophylla*, *E. camaldulensis*, *E. dunnii*, *E. pellita*, *E. maidenii*, *E. saligna* e *E. tereticornis* (Assis, 2000).

Dentre os principais híbridos avaliados para a produção comercial de celulose, destacam-se: (i) *E. urophylla* x *E. grandis* (Bertolucci & Penchel, 1993; Bouvet & Vigneron, 1995; Assis, 2000); (ii) *E. urophylla* x *E. pellita* (Bouvet & Vigneron, 1995; Assis, 2000); (iii) *E. grandis* x *E. tereticornis* (Bouvet & Vigneron,

1995); (iv) *E. urophylla* x *E. globulus* (Assis, 2000); (v) *E. urophylla* x *E. maidenii* (Assis, 2000); (vi) *E. dunnii* x *E. maidenii* (Assis, 2000); (vii) *E. grandis* x *E. maidenii* (Assis, 2000); (viii) *E. grandis* x *E. dunnii* (Assis, 2000); (ix) *E. saligna* x *E. maidenii* (Assis, 2000); (x) *E. saligna* x *E. tereticornis* (Assis, 2000); (xi) *E. nitens* x *E. globulus* (Volker, 1995; Cotterill & Brolin, 1997); (xii) *E. grandis* x *E. smithii*, na África do Sul; (xiii) *E. grandis* x *E. nitens*, na África do Sul. No Brasil, atualmente, ênfase vem sendo dada à hibridação envolvendo o *E. grandis* e o *E. urophylla* com o *E. globulus* e com o *E. smithii*, visando ao melhoramento simultâneo da produtividade e da qualidade da madeira. Também na África do Sul (na indústria Sappi), vem sendo dada ênfase ao melhoramento do híbrido *E. grandis* x *E. smithii*.

Híbridos triplos também têm sido usados com sucesso no melhoramento de *Eucalyptus*, destacando-se os tipos *E. urophylla* x (*E. camaldulensis* x *E. grandis*), *E. urophylla* x (*E. dunnii* x *E. grandis*) e *E. saligna* x (*E. grandis* x *E. urophylla*), os quais, segundo Assis (2000), têm sido mais efetivos na produção de indivíduos para florestas clonais do que os híbridos simples entre estas espécies.

Quanto ao controle genético dos caracteres nos híbridos, resultados preliminares mostram que existe heterose para crescimento (Rezende & Resende, 2000) e não há heterose para os caracteres densidade da madeira, rendimento de celulose, teor de lignina, teor de pentosanas, teor de cinzas e viscosidade, sendo que os híbridos produzem valores intermediários em relação às espécies genitoras (Assis, 2000).

16.9. Heterozigosas e Heteroses Associadas aos Híbridos Multiespécies

A recente utilização de clones comerciais (híbridos e puros) em cruzamentos com outros clones híbridos, visando à seleção de novos clones superiores na descendência, gera um novo cenário no melhoramento de *Eucalyptus*. Este cenário aponta para o melhoramento contínuo de híbridos multiespécies, o qual demanda considerações adicionais sobre as estratégias de melhoramento.

A utilização contínua de clones comerciais híbridos em cruzamentos implica melhoramento contínuo de uma população sintética similar a um composto. A utilização deste esquema apresenta vantagens e riscos. Como vantagens apontam-se: (i) o conhecimento prévio dos genitores para vários caracteres associados à qualidade da madeira, fato que permite um melhor direcionamento dos cruzamentos; (ii) maior probabilidade de obtenção de indivíduos superiores, visto que muitos caracteres desejáveis já se encontram fixados nos genitores; (iii) maior probabilidade de obtenção de recombinantes superiores explorando novos padrões heteróticos (cruzamentos envolvendo maior número de espécies); (iv) possibilidade de uso de maior gama de recursos genéticos. Como riscos pode-se apontar: (i) perda de heterose pela realização de cruzamentos entre clones híbridos de apenas duas espécies; (ii) redução da variabilidade genética

e ocorrência de endogamia, decorrente do exposto em (i); (iii) perda do padrão heterótico envolvendo as várias espécies.

Para evitar os riscos e capitalizar as vantagens, devem ser adotadas estratégias cuidadosas. Neste sentido, serão apresentados alguns aspectos concernentes à utilização de compostos e populações sintéticas.

(a) Híbridos envolvendo duas espécies (*E. urophylla* x *E. grandis*)

Estratégia		Heterozigose (%)*		Heterose	Padrão Heterótico
Seleção recorrente recíproca		100		1 (H _{Ue})	Manutenção
Composto		50		0,5 (H _{UG})	Quebra
Cruzamentos rotacionais		67		0,67 (H _{UG})	Quebra parcial
Geração	Pai	Mãe	Heterozigose (%)*	<i>E. urophylla</i> (%)*	<i>E. grandis</i> (%)*
1	G	U	100	50	50
2	G	$\frac{1}{2}G\frac{1}{2}U$	50	25	75
3	U	$\frac{3}{4}G\frac{1}{4}U$	75	62	38
4	G	$\frac{3}{8}G\frac{5}{8}U$	62,5	31	69
5	U	$\frac{11}{16}G\frac{5}{16}U$	68,75	66	34
6	G	$\frac{11}{32}G\frac{21}{32}U$	65,63		
:	:	:	:	:	:
n	U	$\frac{2}{3}G\frac{1}{3}U$	67	67	33
n + 1	G	$\frac{1}{3}G\frac{2}{3}U$	67	33	67

* % na descendência.

(b) Híbridos envolvendo três espécies [*E. saligna* x (*E. urophylla* x *E. grandis*)]

Estratégia		Heterozigose(%)*		Heterose	Padrão Heterótico
Seleção recorrente recíproca		100		1[(H _{Su} + H _{SG})/2]	Manutenção
Composto		67		0,67[(H _{Su} + H _{SG})/2]	Quebra
Cruzamentos rotacionais		87,5		0,875[(H _{Su} + H _{SG})/2]	Quebra parcial
Gera- ção	Pai	Mãe	Heterozigose (%)*	<i>E. grandis</i> (%)*	<i>E. urophylla</i> (%)* <i>E. saligna</i> (%)*
1	G	U	100	50	50 0
2	S	$\frac{1}{2}G\frac{1}{2}U$	100	25	25 50
3	G	$\frac{1}{4}G\frac{1}{4}U\frac{1}{2}S$	75	62	13 25
4	U	$\frac{5}{8}G\frac{1}{8}U\frac{2}{8}S$	87,5	31	56 13
5	S	$\frac{5}{16}G\frac{9}{16}U\frac{2}{16}S$	87,5	16	28 56

* % na descendência.

(c) Híbridos envolvendo quatro espécies [*E. urophylla* x (*E. grandis* x (*E. globulus* x *E. saligna*))]

Estratégia	Heterozigose(%)*	Heterose
Seleção recorrente recíproca	100	[0,5 H _{UG} + 0,25 H _{UGI} + 0,25 H _{Us}]
Composto	75	0,75 [0,5 H _{UG} + 0,25 H _{UGI} + 0,25 H _{Us}]

* % na descendência.

(d) Híbridos envolvendo cinco espécies [*E. urophylla* x (*E. grandis* x (*E. dunnii* x (*E. globulus* x *E. saligna*)))]

Estratégia	Hetero-zigose (%)*	Heterose
Seleção recorrente recíproca	100	[0,5 H _{UG} + 0,25 H _{UD} + 0,125H _{Ugl} + 0,125 H _{Us}]
Composto	80	0,80 [0,5 H _{UG} + 0,25 H _{UD} + 0,125H _{Ugl} + 0,125 H _{Us}]

* % na descendência.

Para os híbridos entre duas ou três espécies, verifica-se que a seleção recorrente recíproca é a única forma de manter o padrão heterótico entre as populações. A formação de compostos (ou o cruzamento desordenado entre indivíduos híbridos e puros) tende a promover a quebra do padrão heterótico. A utilização de cruzamentos rotacionais é uma boa opção, permitindo a capitalização dos avanços já obtidos nos programas operacionais de plantios e conduzindo a uma quebra apenas parcial do padrão heterótico. É importante destacar que, no híbrido que envolve três espécies, para evitar qualquer quebra no padrão heterótico, é necessária a realização de um programa de seleção recorrente recíproca com uma população híbrida e uma pura, ou dois programas de seleção recorrente recíproca, como por exemplo entre *E. urophylla* x *E. grandis* e *E. urophylla* x *E. globulus* para o melhoramento do híbrido *E. urophylla* x (*E. globulus* x *E. grandis*). Este híbrido triplo contém 50% do germoplasma de *E. urophylla*, 25% do germoplasma de *E. grandis* e 25% do germoplasma de *E. globulus*. Esquema similar pode ser adotado para o melhoramento do híbrido *E. urophylla* x (*E. smithii* x *E. grandis*), o qual contempla 25 % do germoplasma de *E. smithii*. Outra opção é a adoção da SRR modificada (Souza Júnior, 1993) para (*E. grandis* x *E. globulus*) e (*E. urophylla* x *E. globulus*) com SRI para o *E. globulus*. Este esquema envolveria três tipos de progênieis (intrapopulacionais para *E. globulus* e interpopulacionais para os dois tipos de híbrido) e permitiria o uso de híbridos simples de *E. urophylla* e *E. grandis* com *E. globulus* bem como híbridos duplos [(*E. grandis* x *E. globulus*) x (*E. urophylla* x *E. globulus*)] Para os híbridos resultantes de quatro ou cinco espécies, verifica-se que o cruzamento desordenado entre os híbridos (formação de compostos) não é tão prejudicial para a heterozigose, quanto nos híbridos envolvendo menos espécies. Também nestes casos, o uso da seleção recorrente recíproca torna-se praticamente prioritiva (para o melhoramento do híbrido entre quatro espécies seriam necessários

três programas de seleção recorrente recíproca). Estes híbridos podem ser tomados, também, como passos subseqüentes à utilização de híbridos de três espécies por várias gerações, como forma de recuperar e/ou buscar novos padrões heteróticos.

De maneira geral, para os tipos de híbridos descritos de (a) a (d), a utilização dos clones comerciais híbridos e puros, em cruzamento, deve basear-se em cruzamentos rotacionais, sempre escolhendo como material puro a ser cruzado aquela espécie em menor proporção no clone comercial híbrido.

Embora o híbrido "*E. urograndis*" venha sendo usado com sucesso na indústria de celulose de fibra curta, alguns fatores negativos associados à matéria-prima propiciada por este germoplasma podem ser apontados:

- Alto teor de lignina, implicando custos adicionais ao processo de deslignificação e branqueamento da celulose.
- Baixa densidade da madeira, resultando em alto consumo específico de madeira.
- Baixo teor de celulose, redundando em baixo rendimento de celulose.
- Pequena tolerância à seca, ocasionando, na região de plantio, baixa produtividade de matéria-prima em anos com menor pluviosidade.

Visando corrigir as deficiências mencionadas, tem-se a possibilidade de utilização dos seguintes recursos genéticos, com suas respectivas características favoráveis, em termos de valores médios:

E. globulus

Densidade da madeira: 550 kg/m³

Teor de lignina: 21%

Teor de holocelulose: 80%

***E. smithii*:**

Densidade da madeira: 550 kg/m³

Teor de lignina: 20%

Teor de holocelulose: 80%

Estas duas espécies apresentam valores médios similares para as características enumeradas, valores estes mais favoráveis que os apresentados por *E. grandis* e *E. urophylla*. Entretanto, tais espécies apresentam menor potencial de adaptação às condições ambientais das áreas de plantio em regiões tropicais, visto que o *E. globulus* é uma espécie originária de clima temperado e *E. smithii*, de clima subtropical (com área de distribuição natural coincidente, em parte, com a do *E. saligna*). Assim, recomenda-se a avaliação de populações dessas espécies, sobretudo as de baixa altitude, visando à seleção de indivíduos a serem utilizados em programas de hibridação com *E. urophylla* e *E. grandis*. Vale ressaltar que as procedências de *E. globulus* do sul da Tasmânia e sul de Victória (e Bass Strait) têm apresentado crescimento superior às proce-

dências de Portugal e Espanha. Quanto ao *E. smithii*, as procedências Wombeyan Road, Larry's Mountain e Narooma têm apresentado maior crescimento em regiões mais tropicais (Swain et al. 2000).

Para a obtenção de materiais mais tolerantes à seca indica-se a utilização de germoplasma de *E. camaldulensis* (e *E. tereticornis*, que é uma espécie muito próxima ao *E. camaldulensis*) em hibridação com *E. grandis* e *E. urophylla*. Adicionalmente, tais híbridos devem concorrer para o aumento da densidade básica da madeira dos híbridos. Híbridos obtidos destas quatro espécies, com os objetivos de obtenção de maiores tolerância à seca e densidade da madeira, têm sido desenvolvidos com sucesso, especialmente na África do Sul e na Índia.

Podria ser usado também germoplasma de *E. pellita* (espécie muito próxima ao *E. urophylla*) em hibridação com o *E. grandis*, visto que existem no Brasil bons genótipos desta espécie, selecionados em testes de progêniens. O *E. pellita* vem sendo utilizado com sucesso em hibridações no Congo.

16.10. Estratégias de Seleção Recorrente em Função de Aspectos Biológicos das Espécies

As estratégias de melhoramento adotadas em plantas perenes apresentam estreita relação com os sistemas de propagação e com a utilização ou não da heterose. A seguir é apresentada uma classificação genérica dessas estratégias.

(a) Seleção Recorrente Intrapopulacional

a.1) Utilização de sementes de polinização aberta para plantios comerciais

Esta estratégia é utilizada em espécies que não permitem uma eficiente propagação vegetativa e/ou que não apresentam considerável heterose ou capacidade específica de combinação. As espécies podem não apresentar heterose devido à ausência de dominância alélica ou ausência de divergência genética entre as populações. Dentre as espécies em que se utiliza esta estratégia destacam-se: *Pinus taeda*, *Pinus elliotii*, acácia-negra, erva-mate, pupunha, cupuaçu, açaí.

a.2) Utilização de sementes de polinização controlada para plantios comerciais

Esta estratégia tem sido aplicada em espécies com as mesmas características descritas em a.2), porém com facilidade de realização de polinização controlada. Neste caso, a grande vantagem refere-se à maximização da intensidade de seleção. Dentre as espécies em que se utiliza esta estratégia destacam-se o *Pinus radiata* e o *E. globulus*.

a.3) Utilização de clones para plantios comerciais

Esta estratégia é utilizada em espécies que permitem a propagação vegetativa e não apresentam considerável heterose. Tem sido utilizada para o *E. grandis* no Estado de São Paulo, em seringueira, em cafeiro-robusta. Este procedimento também maximiza a intensidade de seleção e permite obter homogeneidade dos produtos. Em espécies alógamas nas quais o interesse reside nos frutos, pode ser necessário o plantio de vários clones compatíveis e não apenas um.

a.4) Utilização de linhagens para plantios comerciais

Este procedimento é empregado em espécies autógamas como o cafeiro-arábica.

(b) Seleção Recorrente Recíproca entre Populações Divergentes

b.1) Utilização de sementes de polinização controlada para plantios comerciais

Esta estratégia é empregada nas espécies com as seguintes características: considerável heterose dada por dominância alélica e divergência genética entre populações; ausência de propagação vegetativa eficiente; facilidade de polinização controlada. Tem sido utilizada em dendezeiro, coqueiro e cacau. Nestas espécies, em geral, são usadas linhagens S₁ (uma geração de autofecundação) para obtenção dos híbridos comerciais, visto que a quantidade de material reprodutivo (pólen e sementes) produzido pelos genitores não endógamos não é suficiente. Nas linhagens S₁ são utilizados vários indivíduos para a produção de pólen e semente.

b.2) Utilização de clones para plantios comerciais

Este procedimento é utilizado em espécies que propiciam eficiente propagação vegetativa e que apresentam considerável heterose para os caracteres de interesse. Tem sido utilizado para o melhoramento do híbrido *E. urophylla* x *E. grandis*.

(c) Seleção Recorrente Recíproca Individual

Neste procedimento, a seleção recorrente recíproca é conduzida entre apenas dois genitores com alta capacidade específica de combinação. Estes genitores são autofecundados gerando duas populações S₁, que são melhoradas, uma em função da outra. Este esquema tem sido empregado no melhoramento do dendezeiro e do eucalipto.

(d) Seleção Recorrente Recíproca em Espécies com Presença de Apomixia

Espécies que apresentam indivíduos apomíticos e também sexuados são comuns dentre as plantas forrageiras como o capim-colonião (*Panicum maximum*) e a braquiária (*Brachiaria* spp.). Tais espécies têm apresentado hetero-

se no cruzamento entre plantas apomíticas (fornecedoras de pólen) e sexuais (utilizadas como genitores femininos). Neste caso, a estratégia ideal de melhoramento é a seleção recorrente recíproca. No entanto, não é possível a recombinação na população apomítica. Dessa forma, a população sexual deve ser melhorada em função da população apomítica, ou seja, dos indivíduos apomíticos em cruzamento. Assim, a recombinação ocorrerá apenas em uma das populações.

Outra opção é a utilização da estratégia relatada em c), porém, com recombinação em apenas uma população S_1 .

Capítulo 17

**Espécies Autógamas,
Híbridos de Linhagens e Progêneres S₁**

17.1. Introdução

Espécies perenes autógamas de importância econômica não existem em grande número. As existentes são de exacerbada importância ao homem, destacando-se o café-arábica, o pessegueiro, o coqueiro-anão e a leguminosa *Leucaena leucocephala*, a qual é reconhecida como uma das mais importantes espécies do grupo das leguminosas arbóreas.

Nas espécies em que ocorre a endogamia, o modelo genético difere daquele modelo simplificado apresentado para as espécies alógamas e, como consequência, as estratégias de melhoramento para essas espécies podem diferir daquelas estudadas e utilizadas para a maioria das espécies alógamas. Neste capítulo são apresentados aspectos relacionados com a genética quantitativa de espécies autógamas perenes e com o uso de progêniest de autofecundação em espécies alógamas. Os modelos genéticos e parâmetros associados apresentados são válidos para espécies diplóides e alopoliplóides (como o café-arábica e a *Leucaena leucocephala*). Em termos citogenéticos e reprodutivos, o comportamento de indivíduos alopoliplóides é similar àquele apresentado por indivíduos diplóides. Isto ocorre porque os genomas em plantas alopoliplóides são muito diferentes, de forma que são formados apenas bivalentes de cromossomos homólogos do mesmo genoma. Assim, em termos de modelos, em genética quantitativa, espécies alopoliplóides podem ser tratadas como diplóides (Wricke & Weber, 1986).

17.2. Melhoramento de Espécies Autógamas Perenes

17.2.1. Modelos genéticos e componentes de variação genotípica

Neste tópico são considerados dois modelos genéticos, baseados nos efeitos aditivo e de dominância, ignorando, portanto, a epistasia. O primeiro modelo, proposto por Cockerham (1954), é adequado a situações em que apenas duas linhagens (cultivares) são cruzadas, objetivando gerar variabilidade para a seleção. O segundo modelo (Cockerham, 1983) é mais geral e apropriado, também, para situações em que várias linhagens são intercruzadas e a seleção é realizada na população segregante. Nestes modelos genéticos, a variabilidade genética total (σ_g^2) em uma geração de endogamia equivale:

Modelo 1. (Cockerham, 1954; Cockerham, 1963)

$$\sigma_g^2 = (1+F)\sigma_a^2 + (1-F^2)\sigma_d^2 \quad (1)$$

Modelo 2. (Cockerham, 1983)

$$\sigma_g^2 = (1+F)\sigma_a^2 + (1-F)\sigma_d^2 + 4FD_1 + FD_2 + F(1-F)H \quad (2)$$

em que:

F : coeficiente de endogamia da geração de autofecundação considerada, associado ao número de gerações de autofecundação a partir dos cruzamentos iniciais.

σ_a^2 : variância genética aditiva.

σ_d^2 : variância genética de dominância.

D_1 : covariância entre os efeitos aditivos dos alelos e os efeitos de dominância dos homozigotos.

D_2 : variâncias dos efeitos de dominância dos homozigotos.

H : soma de quadrados da depressão endogâmica de cada loco.

Pelo modelo 2, com $F = 1$, $\sigma_g^2 = 2\sigma_a^2 + 4D_1 + D_2$.

É importante relatar que na situação de cruzamentos entre apenas duas linhagens, em geral considera-se $p = q = 0,5$ (p e q são as freqüências dos alelos favoráveis e desfavoráveis na população resultante). Neste caso, $D_1 = D_2 = 0$ e $\sigma_d^2 = H$, de forma que se for empregada a expressão 2, obtém-se a expressão 1. No Brasil, a abordagem quantitativa aplicada ao melhoramento de espécies autógamas foi realizada em detalhes por Ramalho & Vencovsky (1978) e Ramalho et al. (1993), considerando o modelo 1, e por Souza Junior (1989) que enfatizou o Modelo 2.

17.2.2. Covariância genética entre parentes

No modelo 2, a covariância genética entre parentes é expressa com função linear dos diversos componentes da variância genética. Assim, para a obtenção das expressões dos progressos genéticos, determina-se a covariância genética [$\text{cov}_g(x,y)$] entre a unidade de seleção (x) e um indivíduo da população melhorada (y), conforme Souza Junior (1989):

$$\text{cov}_g(x,y) = 2r_{xy} \sigma_a^2 + \mu_{xy} \sigma_d^2 + 2(S_{xy} + S_{yx}) D_1 + t_{xy} D_2 + (t_{xy} + V_{xy} - F_x F_y) H$$

em que:

r_{xy} : probabilidade de um alelo, tomado ao acaso em x , ser idêntico por descendência a um outro alelo tomado ao acaso em y : coeficiente de coancestria.

μ_{xy} : probabilidade de dois alelos de x serem idênticos a dois alelos de y e simultaneamente os dois de x e de y serem distintos.

S_{xy} : probabilidade de os dois alelos de x serem idênticos (x é endogâmico), e estes serem idênticos a um alelo tomado ao acaso em y .

S_{yx} : probabilidade de os dois alelos de y serem idênticos (y é endogâmico), e estes serem idênticos a um alelo tomado ao acaso em x .

t_{xy} : probabilidade de os quatro alelos de x e de y serem idênticos por descendência.

V_{xy} : probabilidade de os dois alelos de x serem idênticos por descendência, de os dois alelos de y serem idênticos por descendência, mas de os alelos de x não serem idênticos por descendência aos de y.

F_x e F_y : coeficientes de endogamia dos indivíduos x e y, respectivamente.

17.2.3.Comparação entre processos seletivos

Para comparação entre processos seletivos em espécies autógamas perenes, Resende & Medrado (1994) derivaram as covariâncias entre as unidades de seleção e a população melhorada para situações práticas diferentes daquelas apresentadas por Ramalho & Vencovsky (1978), Souza Júnior (1989) e Vencovsky & Barriga (1992). As expressões apresentadas por tais autores são adequadas apenas a espécies autógamas anuais que não apresentam possibilidades reais de propagação vegetativa, para uso em plantios comerciais e/ou para fixação de genótipos na geração de seleção.

Na obtenção das variâncias e covariâncias foram empregadas as seguintes expressões, segundo o modelo 2:

a) Variância genética total

$$C_{ggg} = \sigma_g^2 = (1 + F_g) \sigma_a^2 + (1 - F_g) \sigma_d^2 + 4 F_g D_1 + F_g D_2 + F_g (1 - F_g) H$$

b) Variância genética entre progênies

$$C_{tgg} = (1 + F_t) \sigma_a^2 + 2(F_g + F_t) D_1 + \frac{(1 - F_g)^2}{(1 - F_t)} \sigma_d^2 + [F_t + \frac{(F_g - F_t)^2}{2(1 - F_t)}] D_2 + \frac{F_t (1 - F_g)^2}{(1 - F_t)} H$$

c) Covariância entre a geração de seleção e a população melhorada totalmente endógama: seleção entre progênies - plantio via sementes ou estacas clonais da geração ∞ .

$$C_{tg\infty} = (1 + F_t) \sigma_a^2 + (F_g + 1 + 2F_t) D_1 + [(F_g + F_t)/2] D_2$$

d) Covariância entre a geração de seleção e a imediatamente posterior: seleção entre progênies - plantio via sementes na geração g' .

$$C_{tgg'} = (1 + F_t) \sigma_a^2 + (F_g + F_{g'} + 2F_t) D_1 + \frac{(1 - F_g)(1 - F_{g'})}{(1 - F_t)} \sigma_d^2 + [F_t + \frac{(F_g - F_t)(F_{g'} - F_t)}{2(1 - F_t)}] D_2 + \frac{F_t + (1 - F_g)(1 - F_{g'})}{1 - F_t} H$$

e) Variância genética dentro de progênies (σ_{gd}^2)

$$\sigma_{gd}^2 = C_{ggg} - C_{tgg}$$

- f) Covariância para o ganho com seleção massal na geração g e plantio da geração g'.**

$$C_{ggg'} = (1 + F_g) \sigma_a^2 + (1 - F_g) \sigma_d^2 + (F_g + 3F_g) D_1 + F_g D_2 + F_g (1 - F_g) H$$

- g) Covariância para o ganho com seleção massal na geração g e plantio da geração ∞ .**

$$C_{gg\infty} = (1 + F_g) \sigma_a^2 + (1 + 3F_g) D_1 + F_g D_2$$

- h) Covariância para o ganho com seleção dentro de progênies na geração g e plantio na geração g'.**

$$CGD_g = C_{ggg'} - C_{tgg'}$$

- i) Covariância para o ganho com seleção dentro de progênies na geração g e plantio na geração ∞ .**

$$CGD\infty = C_{gg\infty} - C_{tg\infty}$$

- j) Covariância para o ganho com seleção massal na geração g e plantio clonal de indivíduos da própria geração g.**

$$CMC = \sigma_g^2 = C_{ggg}$$

- k) Covariância para o ganho com seleção entre progênies na geração g e plantio clonal de progênies da mesma geração.**

$$CEC = C_{tgg}$$

- l) Covariância para o ganho com seleção dentro de progênies na geração g e plantio clonal de indivíduos na mesma geração.**

$$CDC = \sigma_{gd}^2$$

Os índices g, t e g' referem-se à geração atual, à geração precedente a g e a qualquer geração posterior a g, respectivamente.

Na comparação de progressos genéticos, foram considerados caracteres com diferentes padrões de herança quanto ao grau médio de dominância e tipo

de ação gênica, a saber: aditiva, dominância parcial, dominância completa e sobredominância.

17.2.4. Resultados comparativos entre métodos de melhoramento genético de espécies autógamas perenes

O melhoramento genético de espécies autógamas anuais e que não possuem ainda técnicas de propagação vegetativa dominadas fundamenta-se no cruzamento entre diferentes linhagens, para a geração de variabilidade genética adicional, e posterior seleção de novas linhagens, após várias gerações de autofecundação. Assim, as cultivares obtidas sempre representam genótipos totalmente homozigotos e, portanto, não apresentam qualquer variabilidade dentro delas.

As estratégias de melhoramento genético das espécies autógamas perenes deverão diferir das estratégias padrões empregadas no melhoramento de espécies autógamas anuais. Isto ocorre devido à possibilidade de fixação de genótipos superiores nas autógamas perenes em qualquer geração, por meio da propagação vegetativa destes, ou propagação via sementes utilizando-se dos genitores permanentes. Em espécies autógamas anuais (que geralmente não permitem a propagação vegetativa), obrigatoriamente, deve-se caminhar à homozigose total, como única forma de fixação de genótipos superiores.

Em espécies autógamas perenes, os métodos ou estratégias de melhoramento podem ser comparados em função de dois fatores principais: (i) diferentes estratégias de propagação (vegetativa em qualquer geração de autofecundação; via sementes em qualquer geração de autofecundação; via sementes na situação de linhagens totalmente endógamas) e (ii) geração ideal para seleção. Logicamente, esta comparação deve basear-se nos progressos genéticos esperados com seleção.

Fixando a geração de seleção, os métodos de propagação podem ser comparados pela expressão:

$$G_s = k \frac{Cov(US, PM)}{(Var(US))^{1/2}}$$

em que: k é o diferencial de seleção padronizado, $Cov(US, PM)$ é a covariância entre a unidade de seleção e a população melhorada e $Var(US)$ é a variância da unidade de seleção. Nesta situação (para uma geração de autofecundação específica), k e $Var(US)$ são constantes para os diferentes métodos de propagação. Dessa forma, as comparações podem ser baseadas apenas em $Cov(US, PM)$. Essas covariâncias derivadas para os métodos de seleção massal, entre progêneres e dentro de progêneres, são apresentados nas Tabelas 17.1 e 17.2.

Tabela 17.1. Covariâncias* unidade de seleção - população melhorada para os métodos: seleção massal (M), seleção entre progênies (E) e seleção dentro de progênies (D), associados aos métodos de propagação via sementes na própria geração de plantio ($g' =$ uma geração após a seleção) e na situação de linhagens totalmente endógamas (∞).

Geração de Seleção	F**	M _{g'}	M _{∞}	E _{g'}	E _{∞}	D _{g'}	D _{∞}
F ₂	0	$\sigma_a^2 + (1/2)\sigma_d^2$	σ_a^2	-	-	-	--
F ₃	1/2	$(3/2)\sigma_a^2 + (3/8)\sigma_d^2$	$(3/2)\sigma_a^2$	$\sigma_a^2 + (1/8)\sigma_d^2$	σ_a^2	$(1/2)\sigma_a^2 + (1/4)\sigma_d^2$	$(1/2)\sigma_a^2$
F ₄	3/4	$(7/4)\sigma_a^2 + (7/32)\sigma_d^2$	$(7/4)\sigma_a^2$	$(3/2)\sigma_a^2 + (3/32)\sigma_d^2$	$(2/3)\sigma_a^2$	$(1/4)\sigma_a^2 + (1/8)\sigma_d^2$	$(1/4)\sigma_a^2$
F ₅	7/8	$(15/8)\sigma_a^2 + (15/128)\sigma_d^2$	$(15/8)\sigma_a^2$	$(7/4)\sigma_a^2 + (7/128)\sigma_d^2$	$(7/4)\sigma_a^2$	$(1/8)\sigma_a^2 + (1/16)\sigma_d^2$	$(1/8)\sigma_a^2$
F _{∞}	1	-	-	-	$2\sigma_a^2$	-	-

* Consideraram-se cruzamentos entre apenas duas linhagens genitoras e grande número de indivíduos por progénie nos experimentos.

** F: Coeficiente de endogamia associado à geração de seleção.

Observa-se com base nas covariâncias que, de maneira geral, a melhor estratégia de propagação é a clonal, seguida daquela via sementes com plantio da geração g' e, por último, a propagação via sementes com plantio das linhagens totalmente endógamas (Tabelas 17.1 e 17.2).

Logicamente, com a seleção realizada na geração infinita não existem diferenças entre os sistemas de propagação. Fica claro, então, que embora a espécie seja autógama, não compensa obter novas linhagens totalmente homozigotas, exceto se ficar definido F _{∞} como a melhor geração para seleção.

Tabela 17.2. Covariâncias* unidade de seleção - população melhorada para os métodos de seleção: massal (M), entre progênies (E) e dentro de progênies (D), associada à propagação vegetativa na geração de seleção.

Geração de Seleção	F**	M	E	D
F ₂	0	$\sigma_a^2 + \sigma_d^2$	-	-
F ₃	1/2	$(3/2)\sigma_a^2 + (3/4)\sigma_d^2$	$\sigma_a^2 + (1/4)\sigma_d^2$	$(1/2)\sigma_a^2 + (1/2)\sigma_d^2$
F ₄	3/4	$(7/4)\sigma_a^2 + (7/16)\sigma_d^2$	$(3/2)\sigma_a^2 + (3/16)\sigma_d^2$	$(1/4)\sigma_a^2 + (1/4)\sigma_d^2$
F ₅	7/8	$(15/8)\sigma_a^2 + (15/64)\sigma_d^2$	$(7/4)\sigma_a^2 + (7/64)\sigma_d^2$	$(1/8)\sigma_a^2 + (1/8)\sigma_d^2$
F _{∞}	1	-	$2\sigma_a^2$	-

* Foram considerados os cruzamentos entre apenas duas linhagens genitoras e grande número de indivíduos por progénie nos experimentos.

** F : Coeficiente de endogamia associado à geração de seleção.

É importante ressaltar que as inferências realizadas são válidas também para o método do índice multiefeitos, cujo valor de Cov (US, PM) é função desses valores para os métodos da seleção entre progêniess e dentro de progêniess.

A definição da geração ideal de seleção é dependente da herança do caráter considerado, principalmente quanto ao grau médio de dominância. Considerando, inicialmente, os graus médio de dominância (g.m.d.) inerentes à dominância completa (g.m.d. = 1) e sobredominância (g.m.d. > 1), pode-se afirmar que a seleção deve ser praticada na geração F_1 , pois com seleção em gerações de autofecundação (F_2 por exemplo), o máximo que se consegue (selecionando o melhor indivíduo) é um indivíduo com valor genético/genotípico idêntico ao de F_1 . Assim, para casos de sobredominância e dominância completa, a estratégia de melhoramento deve restringir-se à realização de cruzamentos dialélicos ou fatoriais, entre genitores selecionados, com posterior seleção e propagação clonal do melhor híbrido. Entretanto, para geração de variabilidade genética adicional para futuras seleções, um esquema de seleção recorrente é recomendável, preferencialmente a seleção recíproca (Resende & Higa, 1990; Ramalho et al. 1999).

Com dominância parcial e ausência de dominância (herança totalmente aditiva), existe certa superioridade de genótipos homozigotos sobre heterozigotos e a condução de gerações de autofecundação pode ser uma boa opção. Nesta situação, a definição da geração ideal de seleção deve basear-se na estimativa de ganhos genéticos com seleção.

Com herança completamente aditiva, não existe depressão endogâmica e as médias das diferentes gerações não se alteram. Considerando que a variância fenotípica pouco varia entre as gerações, depreende-se que as covariâncias unidades de seleção - população melhorada (Tabelas 17.1 e 17.2) são preponderantes na comparação dos ganhos. É imperativo dizer que, nesse caso, as covariâncias apresentadas são função apenas da variância genética aditiva (substituindo σ_d^2 por 0 nas Tabelas 17.1 e 17.2) e, portanto, não existem diferenças quanto aos três sistemas de propagação. Verifica-se, assim, que nesse caso é vantajoso trabalhar com linhagens completamente endogâmicas e selecionar em F_∞ , pois Cov (US,PM) = 2 σ_a^2 contra Cov (US,PM) = σ_a^2 da geração F_2 .

Se existir algum nível de dominância alélica, existe depressão endogâmica e, portanto, as médias das diferentes gerações não serão as mesmas. Assim, para o caso de dominância parcial, a determinação da geração ideal de seleção deverá levar em conta também as diferentes médias das gerações. A média de cada geração de autofecundação é dada por $M_1 = M_0 - 2F \sum p q d$ (Falconer, 1989) em que d é o valor genotípico do heterozigoto, M_0 a média original e \sum indica somatório sobre todos os locos. Assim, a redução na média é diretamente proporcional ao coeficiente de endogamia, dependendo também das freqüências alélicas, do nível de dominância e do número de locos controlando o

caráter. Com dominância parcial, d varia no intervalo aberto de zero a um, de forma que a depressão será mais baixa e alta para valores de d próximos de zero e um, respectivamente.

Analisando inicialmente as covariâncias e sabendo que para o grau médio de dominância 1,0 e $p = q = 0,5$, $\sigma_d^2 = 1/2 \sigma_a^2$ (Vencovsky & Barriga, 1992), verifica-se que elas aumentam com as gerações de autofecundação (Tabelas 17.1 e 17.2). Como as médias populacionais decrescem com as autofecundações, não fica claro, para casos de dominância parcial, qual a melhor geração para se efetuar a seleção, visando à propagação vegetativa ou via sementes com fixação dos genótipos na geração de seleção. Com dominância baixa, a geração ideal de seleção tenderá para F_∞ , conforme verificado para os casos de herança totalmente aditiva. Com dominância alta, a geração ideal de seleção tenderá para F_1 ou F_2 (S_0), segundo o verificado nos casos de dominância completa ou sobredominância. Logicamente, com dominância intermediária, uma geração com F intermediário (F_3 , F_4) deverá ser a ideal.

Torna-se necessário, então, um conhecimento prévio da herança do caráter, para determinar com segurança a geração ideal para seleção. Em *Leucaena leucocephala* (Brewbaker, 1985) e café-arábica (Cilas et al., 1998a), foi mostrada a existência de heterose para vários caracteres de interesse, o que é um indício da existência de dominância alélica. Assim sendo, pode-se dizer que, para o melhoramento desses caracteres, dificilmente será vantajoso retardar a seleção para gerações avançadas. Pelo contrário, realizar a seleção logo após a hibridação (F_1) ou em F_2 , associada à propagação vegetativa, deverá ser a estratégia mais eficiente. Em pessegueiro a estratégia de hibridação seguida de seleção e propagação vegetativa já vem sendo usada rotineiramente (Bruckner, 1999).

17.2.5. Estratégias de utilização do material genético melhorado

Sendo uma espécie autógama, o plantio de uma única linhagem de *Leucaena leucocephala* conduz a povoamentos bastante homogêneos. Baseando-se neste fato, Brewbaker (1985) descreve as seguintes soluções genéticas para a problemática: plantio de várias linhagens simultaneamente e uso de variedades multilinhas.

Quanto ao plantio de várias linhagens, o autor mostra que pelo menos três linhagens distintas deveriam ser utilizadas em qualquer plantação de grande porte. Com respeito ao tipo Salvadorenho (arbóreas), as linhagens K8, K28, K29, K67 e K132 não diferem significativamente na produção de madeira e forragem (Brewbaker, 1985) e poderiam ser utilizadas para estes plantios mistos.

O termo multilinhas designa mistura de linhagens de espécies autógamas e, segundo Brewbaker (1985), esta mistura poderia ser obtida em *Leucaena leucocephala* por meio de cruzamentos entre várias linhagens e utilização da geração F_2 para plantio.

Uma terceira estratégia pode ser proposta: a utilização de cultivares predominantemente heterozigotos por meio do emprego da propagação vegetativa ou via sementes com fixação de genótipos na geração de seleção. Esta estratégia, além de garantir maior heterogeneidade nos plantios, é adequada em termos de maior eficiência no processo seletivo para caracteres que apresentam dominância alélica. Esta estratégia vem sendo utilizada para o café-robusta (espécie alógama), em que pelo menos cinco clones distintos são plantados simultaneamente (Ferrão et al., 1999).

A técnica de propagação vegetativa já vem sendo desenvolvida para *Leucaena leucocephala*, que é classificada como uma espécie de fácil enraizamento (Puri & Rajora, 1992). Em café-arábica, o desenvolvimento da tecnologia de propagação vegetativa é, atualmente, uma das prioridades de pesquisa (Sera, 1998; Mendes, 1999; Sakiyama et al., 1999). Em café-robusta (espécie alógama), a propagação vegetativa já é realidade, e plantios comerciais de clones vêm sendo realizados com sucesso no Estado do Espírito Santo. Em pessegueiro, a propagação vegetativa é usada rotineiramente (Bruckner, 1999).

17.2.6. Estratégias para coleta de germoplasma, experimentação e melhoramento no longo prazo

Na coleta de germoplasma, deve-se ter em mente que a quase totalidade da variabilidade genética existente na espécie encontra-se entre as populações e não dentro das populações. O contrário ocorre nas espécies alógamas. Assim, a estratégia de coleta de germoplasma deve enfatizar a coleta de várias (todas, se possível) populações, amostrando-se um ou poucos indivíduos por população. Apenas um indivíduo por região amostrada é suficiente, quando não coexistem várias linhagens em um só local. A quantificação da representatividade genética de amostras de plantas e/ou sementes de espécies autógamas deve basear-se nas expressões do tamanho efetivo populacional para tais espécies, conforme apresentado no tópico 18.7.

De posse do germoplasma base, a estratégia inicial óbvia é a comparação entre as diferentes linhagens, que objetivam à seleção e à utilização das melhores como fonte de sementes para plantios comerciais. A experimentação, neste caso, deve procurar maximizar o número de repetições em detrimento do número de plantas por parcela, já que não existe variabilidade dentro dos acessos. O uso de parcelas pequenas tem levado, na prática, a baixos coeficientes de variação experimental, fornecendo, portanto, a precisão necessária (Brewbaker, 1983). Em café-arábica, parcelas com uma planta têm sido utilizadas com sucesso (Cilas et al., 1998a).

Entretanto, o ponto principal do melhoramento da espécie deve ser a geração de nova variabilidade e a consequente obtenção de genótipos superiores àqueles existentes em condições naturais ou na geração atual. Esta estratégia poderia contemplar o cruzamento entre as melhores linhagens existentes, em esquema de dialélico incompleto (sem os genitores e recíprocos) e posterior

seleção dos genótipos desejados na descendência. Na seleção dos genitores (linhagens) a serem cruzados, deve-se enfatizar o seu comportamento médio, bem como a divergência genética entre eles. Este procedimento concorre para a maior probabilidade de obtenção de genótipos transgressivos nas gerações segregantes. Nesta etapa, a aplicação de técnicas multivariadas para estimação da divergência genética é recomendável (Resende & Higa, 1991; Cruz & Regazzi, 1994). A maneira como deverá ser realizada a seleção, a partir da hibridação, foi discutida em tópico anterior.

Para o melhoramento no longo prazo em espécies perenes autógamas como o café-arábica, Ramalho et al. (1999) sugeriram o emprego da seleção recorrente, como forma de gerar linhagens superiores via aumento da freqüência dos alelos favoráveis nas populações. Tais autores recomendam, preferencialmente, a seleção nas gerações iniciais (F_3), enfatizando que, em virtude do D_1 poder assumir valores negativos, o fato de postergar a seleção para gerações mais avançadas pode reduzir o ganho esperado com a seleção.

Em resumo, pode-se concluir que:

- Dentre as estratégias de propagação comparadas, fixando a geração de seleção (exceto F_∞), concorrem para maior ganho genético, em ordem decrescente: vegetativa; via sementes fixando os genótipos superiores na geração de seleção; via sementes utilizando linhagens totalmente endógamas.
- A geração de autofecundação ideal para seleção é dependente do grau médio de dominância do caráter sob melhoramento. Com dominância e sobredominância, a seleção deverá ser realizada em F_1 . Com herança completamente aditiva, a seleção deve ser praticada em F_∞ . Com dominância parcial, a geração ideal de seleção tenderá a: F_∞ para graus médio de dominância no intervalo de 0 a 0,5; F_1 e F_2 para graus de dominância no intervalo 0,5 a 1,0.
- Para caracteres produtivos, os quais parecem exibir dominância, a estratégia ideal seria a seleção nas gerações F_1 ou F_2 , com propagação vegetativa do material selecionado.

17.3. Híbridos de Linhagens

17.3.1. Estimação de parâmetros genéticos a partir de cruzamentos entre várias linhagens

Os delineamentos (Capítulo 1) de cruzamento empregados na estimação de parâmetros genéticos em espécies alógamas podem, também, ser aplicados em espécies autógamas. Neste caso, a composição da covariância genética entre parentes muda, visto que o coeficiente de endogamia (F) dos genitores utilizados não é 0, como nas espécies alógamas mas, em geral, $F = 1$ (cruzamento entre linhagens homozigotas).

(a) cruzamentos biparentais

Para o delineamento de cruzamentos biparentais, progênieis de irmãos germanos são geradas e a covariância genética entre estes irmãos é dada por:

$$COV(IG) = \frac{1+F}{2} \sigma_a^2 + \left(\frac{1+F}{2}\right)^2 \sigma_d^2 = \sigma_a^2 + \sigma_d^2, \text{ quando } F = 1$$

O esquema da análise de variância considerando o delineamento em blocos casualizados (b blocos) e parcelas com uma planta equivale a:

FV	QM	E(QM)
Blocos	-	-
Progênieis (cruzamentos)	Q ₁	$\sigma^2 + b\sigma_p^2$
Resíduo	Q ₂	σ^2

Com F=1, têm-se, então, os seguintes estimadores:

$$\hat{\sigma}_p^2 = \frac{Q_1 - Q_2}{b} = \hat{\sigma}_a^2 + \hat{\sigma}_d^2 : \text{variância entre progênieis.}$$

$$\hat{\sigma}^2 = Q_2 : \text{variância residual (ambiental).}$$

$$\hat{h}_g^2 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}^2} = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_a^2 + \sigma_d^2 + \sigma^2} : \text{herdabilidade individual no sentido amplo.}$$

Este delineamento não permite estimar a herdabilidade no sentido restrito. No caso de parcelas com várias (n) plantas, tem-se, adicionalmente, a fonte de variação dentro de parcelas, com variância σ_e^2 , puramente ambiental, e $\sigma^2 = \frac{\sigma_d^2}{n} + \sigma_e^2$, em que σ_e^2 é a variância entre parcelas.

(b) cruzamentos hierárquicos

Neste delineamento, genitores femininos são cruzados (m vezes) com grupos desconexos de machos. Neste caso, o esquema da análise de variância, considerando o delineamento em blocos casualizados (b blocos) e parcelas com uma planta, equivale a:

FV	QM	E(QM)
Blocos	-	-
Fêmeas	Q ₁	$\sigma^2 + b\sigma_{C/F}^2 + bm\sigma_f^2$
Machos/Fêmeas (cruzamento/F)	Q ₂	$\sigma^2 + b\sigma_{C/F}^2$
Indivíduo/Macho/Fêmea	Q ₃	σ^2

Com $F = 1$, tem-se:

$$COV(IG) = \sigma_a^2 + \sigma_d^2 = \sigma_{C/F}^2 + \sigma_f^2$$

$$COV(MI) = \frac{1+F}{4} \sigma_a^2 = (1/2) \sigma_a^2 = \sigma_f^2$$

$$\sigma_{C/F}^2 = COV(IG) - COV(MI) = (1/2) \sigma_a^2 + \sigma_d^2$$

E os estimadores são:

$$\hat{\sigma}_f^2 = \frac{Q_1 - Q_2}{bm} = (1/2) \hat{\sigma}_a^2 : \text{variância entre fêmeas.}$$

$$\hat{\sigma}_{C/F}^2 = \frac{Q_2 - Q_3}{b} = (1/2) \hat{\sigma}_a^2 + \hat{\sigma}_d^2 : \text{variância entre cruzamentos dentro de fêmeas.}$$

$$\hat{\sigma}^2 = Q_3 : \text{variância residual (ambiental).}$$

$$\hat{\sigma}_a^2 = 2\hat{\sigma}_f^2 : \text{variância genética aditiva.}$$

$$\hat{\sigma}_d^2 = \hat{\sigma}_{C/F}^2 - \hat{\sigma}_f^2 : \text{variância genética de dominância.}$$

$$\hat{h}_a^2 = \frac{2 \hat{\sigma}_f^2}{\hat{\sigma}_f^2 + \hat{\sigma}_{C/F}^2 + \hat{\sigma}^2} = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}^2} : \text{herdabilidade individual no sentido restrito.}$$

$$\hat{h}_g^2 = \frac{\hat{\sigma}_f^2 + \hat{\sigma}_{C/F}^2}{\hat{\sigma}_f^2 + \hat{\sigma}_{C/F}^2 + \hat{\sigma}^2} = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_d^2}{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}^2} : \text{herdabilidade individual no sentido amplo.}$$

(c) Cruzamentos fatoriais

Cruzando m machos com f fêmeas, o esquema da análise de variância, considerando o delineamento de blocos ao acaso (b blocos) e parcelas de uma planta, é:

FV	QM	E(QM)
Blocos	-	-
Machos (M)	Q_1	$\sigma^2 + b\sigma_{fm}^2 + bf\sigma_m^2$
Fêmeas (F)	Q_2	$\sigma^2 + b\sigma_{fm}^2 + bm\sigma_f^2$
$F \times M$	Q_3	$\sigma^2 + b\sigma_{fm}^2$
Indivíduo/ ($F \times M$)	Q_4	σ^2

Com $F = 1$, tem-se:

– Covariância de meios-irmãos maternos

$$COV(MIM) = \sigma_f^2 = (1/2) \sigma_a^2$$

– Covariância de meios-irmãos paternos

$$COV(MIP) = \sigma_m^2 = (1/2) \sigma_a^2$$

– Covariância entre irmãos germanos

$$COV(IG) = \sigma_a^2 + \sigma_d^2 = \sigma_f^2 + \sigma_m^2 + \sigma_{fm}^2 = (1/2) \sigma_a^2 + (1/2) \sigma_a^2 + \sigma_d^2$$

– Variância de dominância

$$\sigma_d^2 = COV(IG) - COV(MIM) - COV(MIP) = \sigma_{fm}^2$$

Os estimadores são:

$$\hat{\sigma}_m^2 = \frac{Q_1 - Q_3}{bf} = (1/2) \hat{\sigma}_d^2 : \text{variância entre machos.}$$

$$\hat{\sigma}_f^2 = \frac{Q_2 - Q_3}{bm} = (1/2) \hat{\sigma}_d^2 : \text{variância entre fêmeas.}$$

$$\hat{\sigma}_{fm}^2 = \frac{Q_3 - Q_4}{b} = \hat{\sigma}_d^2 : \text{variância da interação fêmea x machos.}$$

$$\hat{\sigma}^2 = Q_4 : \text{variância residual (ambiental).}$$

$$\hat{\sigma}_a^2 = 2\sigma_f^2 \text{ ou } 2\sigma_m^2 : \text{variância genética aditiva.}$$

$$\hat{\sigma}_d^2 = \hat{\sigma}_{fm}^2 : \text{variância genética de dominância.}$$

$$\hat{h}_a^2 = \frac{\hat{\sigma}_f^2 + \hat{\sigma}_m^2}{\hat{\sigma}_f^2 + \hat{\sigma}_m^2 + \hat{\sigma}_{fm}^2 + \hat{\sigma}^2} : \text{herdabilidade individual no sentido restrito.}$$

$$\hat{h}_g^2 = \frac{\hat{\sigma}_f^2 + \hat{\sigma}_m^2 + \hat{\sigma}_{fm}^2}{\hat{\sigma}_f^2 + \hat{\sigma}_m^2 + \hat{\sigma}_{fm}^2 + \hat{\sigma}^2} : \text{herdabilidade individual no sentido amplo.}$$

(d) Dialélico triangular (sem os recíprocos) sem os genitores

Considerando o delineamento de blocos ao acaso (b blocos) e parcelas de uma planta, o esquema de análise será:

FV	QM	E(QM)
Blocos	-	-
CGC	Q ₁	$\sigma^2 + b\sigma_s^2 + b(g-2)\sigma_q^2$
CEC	Q ₂	$\sigma^2 + b\sigma_s^2$
Resíduo	Q ₃	σ^2

Com F = 1, tem-se:

$$COV(IG) = \sigma_a^2 + \sigma_d^2 = 2\sigma_q^2 + \sigma_s^2$$

$$COV(MI) = (1/2) \sigma_a^2 = \sigma_q^2$$

Têm-se, então, os estimadores:

$$\hat{\sigma}_q^2 = \frac{Q_1 - Q_2}{b(g-2)} = (1/2) \hat{\sigma}_d^2 : \text{variância da capacidade geral de combinação (CGC).}$$

$$\hat{\sigma}_s^2 = \frac{Q_2 - Q_3}{b} = \hat{\sigma}_d^2 : \text{variância da capacidade específica de combinação (CEC).}$$

$$\hat{\sigma}^2 = Q_3 : \text{variância residual ambiental.}$$

$$\hat{\sigma}_a^2 = 2\hat{\sigma}_g^2 = \text{variância genética aditiva.}$$

$$\hat{\sigma}_d^2 = COV(IG) - 2COV(MI) = \hat{\sigma}_s^2 : \text{variância genética de dominância.}$$

$$\hat{h}_a^2 = \frac{2 \hat{\sigma}_q^2}{2\hat{\sigma}_q^2 + \hat{\sigma}_s^2 + \hat{\sigma}^2} = \frac{\hat{\sigma}_q^2}{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}^2} : \text{herdabilidade individual no sentido restrito.}$$

$$\hat{h}_g^2 = \frac{2\hat{\sigma}_q^2 + \hat{\sigma}_s^2}{2\hat{\sigma}_q^2 + \hat{\sigma}_s^2 + \hat{\sigma}^2} = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_d^2}{\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}^2} : \text{herdabilidade individual no sentido amplo.}$$

Outro delineamento muito relevante é o delineamento dialélico parcial circulante (Kempthorne & Curnow, 1961; Keuls & Garretsen, 1977), o qual não é muito utilizado na prática, uma vez que os melhoristas, em geral, preferem usar pequenos delineamentos, tais quais os cruzamentos fatoriais desconexos, que são similares. Tal delineamento permite estimar a variância aditiva e de dominância com a mesma precisão (nos demais dialélicos e fatoriais, a variância de dominância é estimada com maior precisão do que a variância aditiva) (Cilas, 1995b). O dialélico circulante apresenta a vantagem de ser um delineamento conexo, o que é desejável em termos de avaliação genética.

Para espécies com sistema reprodutivo misto podem ser adotados os mesmos modelos, bastando substituir $F = S/(2-S)$ nas expressões de $COV(IG)$ e $COV(MI)$.

Como exemplo, considere os resultados da avaliação de cruzamentos entre 10 genitores distintos de café-arábica no delineamento fatorial 5 x 5, com 15 repetições de uma planta por parcela.

FV	QM	E(QM)
Blocos	-	-
Machos (M)	$Q_1=12,78$	$\sigma^2 + 15 \sigma_{fm}^2 + 75 \sigma_m^2$
Fêmeas (F)	$Q_2=13,22$	$\sigma^2 + 15 \sigma_{fm}^2 + 75 \sigma_f^2$
M x F	$Q_3= 1,97$	$\sigma^2 + 15 \sigma_{fm}^2$
Resíduo	$Q_4= 0,71$	σ^2

São as seguintes as estimativas:

$$\hat{\sigma}_f^2 = \frac{Q_2 - Q_3}{bm} = \frac{13,22 - 1,97}{75} = 0,15$$

$$\hat{\sigma}_m^2 = \frac{Q_1 - Q_3}{bf} = \frac{12,78 - 1,97}{75} = 0,1441$$

$$\hat{\sigma}_{mf}^2 = \frac{Q_3 - Q_4}{15} = \frac{1,97 - 0,71}{15} = 0,084$$

$$\hat{\sigma}^2 = Q_4 = 0,71$$

$$\hat{h}_a^2 = \frac{\hat{\sigma}_f^2 + \hat{\sigma}_m^2}{\hat{\sigma}_f^2 + \hat{\sigma}_m^2 + \hat{\sigma}_{mf}^2 + \hat{\sigma}^2} = \frac{0,15 + 0,1441}{0,15 + 0,1441 + 0,084 + 0,71} = 0,2703$$

$$\hat{h}_g^2 = \frac{\hat{\sigma}_f^2 + \hat{\sigma}_m^2 + \hat{\sigma}_{fm}^2}{\hat{\sigma}_f^2 + \hat{\sigma}_m^2 + \hat{\sigma}_{mf}^2 + \hat{\sigma}^2} = \frac{0,15 + 0,1441 + 0,084}{0,15 + 0,1441 + 0,084 + 0,71} = 0,3475$$

17.3.2. BLUP e predição de valores genotípicos por meio dos valores fenotípicos dos próprios híbridos

Para a utilização de híbridos de linhagens, torna-se essencial a predição dos valores genotípicos dos híbridos, para o propósito de seleção. Um procedimento comum e recomendável é a realização de cruzamentos dialélicos ou fatoriais entre as linhagens e posterior seleção dos híbridos. Neste caso, o modelo linear associado a análise apenas dos híbridos (experimentos com uma planta por parcela) é o seguinte:

$$Y_{ijk} = \mu + B_K + G_i + G_j + S_{ij} + E_{ijk}, \text{ em que:}$$

Y_{ijk} : valor fenotípico do cruzamento ij no bloco k (árvore k do cruzamento ij).

G_i : capacidade geral de combinação do genitor feminino i, aleatório.

G_j : capacidade geral de combinação do genitor masculino j, aleatório.

S_{ij} : capacidade específica de combinação do cruzamento ij, aleatório.

B_K : efeito do bloco k, fixo.

E_{ijk} : desvio aleatório dentro do cruzamento ij, aleatório.

A partir deste modelo as herdabilidades individuais no sentido restrito (h_a^2) e no sentido amplo (h_g^2) podem ser estimadas com base nas covariâncias entre parentes (meios-irmãos e irmãos germanos) originários de genitores completamente endogâmicos ($F = 1$), conforme apresentado no tópico anterior.

Para o café-arábica, a partir de um cruzamento dialélico entre sete linhagens genitoras, Cilas et al. (1998a) obtiveram as seguintes estimativas: $h_a^2 = 0,26; 0,001; 0,37$ e $0,04$; $h_g^2 = 0,37; 0,22; 0,49$ e $0,16$, para os caracteres produção acumulada de quatro anos, diâmetro da planta, altura da planta e número de ramos primários, respectivamente.

O valor genotípico dos híbridos (a serem reconstituídos a cada geração e propagados por sementes) é dado por:

$$\hat{H}_{ij} = \hat{a}_{ij} + \hat{d}_{ij},$$

$$= 1/2 \hat{a}_i + 1/2 \hat{a}_j + \hat{d}_{ij}, \text{ em que:}$$

$$\hat{a}_i = \frac{bm h_a^2}{1 + (b-1) h_g^2 + (m-2)b(1/2) h_a^2} (\bar{Y}_{i..} - \bar{Y}_{...}) = \frac{bm h_a^2}{1 + (b-1) h_g^2 + (m-2)b(1/2) h_a^2} \hat{G}_i$$

$$\hat{a}_j = \frac{bf h_a^2}{1 + (b-1) h_g^2 + (f-2)b(1/2) h_a^2} (\bar{Y}_{.j} - \bar{Y}_{...}) = \frac{bf h_a^2}{1 + (b-1) h_g^2 + (f-2)b(1/2) h_a^2} \hat{G}_j$$

$$\hat{d}_{ij} = \frac{1 + (b-1)(h_g^2 - h_a^2)}{1 + (b-1) h_g^2 - b h_a^2} (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...}) = \frac{1 + (b-1)(h_g^2 - h_a^2)}{1 + (b-1) h_g^2 - b h_a^2} \hat{S}_{ij},$$

em que m, f e b referem-se ao número de genitores masculinos, femininos e de repetições, respectivamente.

Preditores para delineamentos com várias plantas por parcela foram apresentados por Resende (1999b).

O valor genotípico dos híbridos pode ser predito também de forma indireta, usando outros caracteres morfológicos por meio do preditor:

$$\hat{H}_{ij} = C P^{-1} \hat{G}_{ij}, \text{ em que:}$$

$$C = [\Sigma(H_{ij}, \hat{G}_i) \quad \Sigma(H_{ij}, \hat{G}_j) \quad \Sigma(H_{ij}, \hat{S}_{ij})]$$

: matriz de covariância entre o verdadeiro valor genotípico para o caráter sendo predito (preditando) e os valores estimados de G_i, G_j e S_{ij} associados aos caracteres preditores (dimensão q x 3, em que q é o número de caracteres preditores).

$$P = \begin{bmatrix} \sum \hat{G}_i \hat{G}_i & \sum \hat{G}_i \hat{G}_j & \sum \hat{G}_i \hat{S}_{ij} \\ \sum \hat{G}_j \hat{G}_i & \sum \hat{G}_j \hat{G}_j & \sum \hat{G}_j \hat{S}_{ij} \\ \sum \hat{S}_{ij} \hat{G}_i & \sum \hat{S}_{ij} \hat{G}_j & \sum \hat{S}_{ij} \hat{S}_{ij} \end{bmatrix}$$

: matriz de variância-covariância envolvendo \hat{G}_i , \hat{G}_j e \hat{S}_{ij} associados aos q caracteres preditores (3q x 3q).

$$\hat{G}_{ij} = \begin{bmatrix} \hat{g}_i \\ \hat{g}_j \\ \hat{S}_{ij} \end{bmatrix} \hat{g}_i : \text{ vetor coluna (q x 1) dos efeitos de } G_i \text{ associados aos q caracteres.}$$

: vetor coluna (3q x 1) dos efeitos de G_i, G_j e S_{ij} associados aos q caracteres.

$$\hat{S}_{ij} : \text{ vetor coluna (q x 1) dos efeitos de } S_{ij} \text{ associados aos q caracteres.}$$

Cilas et al. (1998a) aplicaram este último preditor em café-arábica e concluíram que a produção pode ser perfeitamente predita pela combinação dos caracteres morfológicos altura da planta, diâmetro do caule e número de ramos primários. Um índice combinando estes três caracteres propiciou o mesmo ganho genético esperado com a seleção direta para a produção acumulada de quatro anos. As correlações genéticas com a produção acumulada foram de 0,24; -0,55 e 0,92 para diâmetro, altura e número de ramos primários, respectivamente. As correlações fenotípicas foram de 0,48; 0,08 e 0,60, para diâmetro, altura e número de ramos primários, respectivamente. Adicionalmente, os autores verificaram a presença de significância da capacidade específica de combinação para os quatro caracteres, fato que corrobora a existência de heterose em café-arábica.

Um procedimento mais refinado de predição em plantas perenes é o BLUP (Henderson, 1984), o qual considera adequadamente o desbalanceamento dos dados. Para a situação em que se utilizam as médias dos híbridos (entre linhagens não aparentadas) por repetições, o seguinte modelo pode ser empregado (Resende et al., 1996a).

$$\bar{Y}_{ijk.} = \mu + B_K + H_{ij} + e_{ijk} + \delta_{ijk}/n_{ijk}, \text{ em que:}$$

\bar{Y}_{ijk} : média do híbrido na repetição k.

B_K : efeito do bloco k, fixo.

H_{ij} : efeito do híbrido ij, aleatório.

e_{ijk} : efeito da parcela ijk, aleatório.

δ_{jkl} : efeito de planta dentro de parcela.

n_{ijk} : número de plantas na parcela ijk.

Por este modelo, o BLUP de H_{ij} é dado por:

$$\hat{H} = CV^{-1}(y - X\hat{\beta})$$

$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$, em que:

V : matriz de covariância fenotípica entre as médias (por repetição) dos híbridos. Apresenta em sua diagonal os valores:

$$V(\bar{Y}_{ijk}) = \sigma_H^2 + \sigma_e^2 + \sigma_\delta^2/n_{ijk} = \sigma_H^2 + \sigma^2, \text{ em que } \sigma^2 = \sigma_e^2 + \sigma_\delta^2/n_{ijk}.$$

Apresenta nos elementos fora da diagonal, os valores:

- covariância entre médias de híbridos não oriundos de linhagens comuns:

$$COV(\bar{Y}_{ijk}, Y_{ij'k}) = 0$$

- covariância entre médias de híbridos oriundos de uma linhagem em comum:

$$COV(\bar{Y}_{ijk}, Y_{ij'k}) = (1/2) \sigma_a^2, \text{ considerando que todas as linhagens são não aparentadas.}$$

- covariâncias entre médias de um mesmo híbrido em diferentes repetições: $COV(\bar{Y}_{ijk}, Y_{ij'k}) = \sigma_H^2$.

- C : matriz de covariância entre H_{ij} e \bar{Y}_{ijk} . Seus elementos não-zero equivalem a $\sigma_H^2 = \sigma_a^2 + \sigma_\delta^2$: variância genética entre os híbridos, ou a $(1/2) \sigma_a^2$.

- $\hat{\beta}$: estimador de quadrados mínimos generalizados (GLS) dos efeitos de blocos, o qual propicia as melhores estimativas lineares não viciadas (BLUE) dos efeitos de blocos.

Este mesmo procedimento pode ser aplicado em espécies alógamas com o objetivo de predição dos efeitos genéticos da família (\hat{a}_e). Tal procedimento seguido da predição de valores genéticos referentes aos efeitos de indivíduo dentro de progênie ($\hat{a}_d = b_1(Y_{ijk} - X\hat{\beta} - (1/2) \hat{a}_e)$) e da soma $\hat{a} = \hat{a}_e + \hat{a}_d$ fornece a predição BLUP do indivíduo, conforme usado por Duda et al. (1997). Isto se deve ao fato de os desbalanceamentos de dados ocorrerem ao nível do efeito de progênie e de tal desbalanceamento ser considerado quando da aplicação do BLUP para o efeito de progênie. Neste caso, \bar{Y}_{ijk} referem-se a médias da família

ij (geralmente de meios-irmãos) no bloco k e σ_H^2 deve ser substituído pela variância entre progêneres (σ_p^2). Também, com genitores não parentados $COV(\bar{Y}_{ijk}, Y_{ij'k}) = 0$. Pode-se, também, trabalhar com os dados individuais, porém obtendo-se o BLUP de família.

Um procedimento BLUP mais generalizado, adequado também às situações de linhagens parentadas e híbridos avaliados em vários ensaios é apresentado no tópico seguinte.

17.3.3. BLUP e predição de valores genotípicos de híbridos não avaliados

Em muitos programas de melhoramento, grande número de linhagens existe e não é viável obter e avaliar todos os híbridos possíveis. Procedimentos de predição dos valores genotípicos de híbridos não avaliados em campo têm sido estudados, sendo que a maioria deles não são adequados, tais quais os usos da performance per se das linhagens e da diversidade genética molecular (marcadores) entre as linhagens genitoras não parentadas (Bernardo, 1999). Em café-arábica, Cilas et al. (1998a) mostraram que não existe correlação entre a performance da linhagem e a sua capacidade geral de hibridação.

Aparentemente, o único procedimento adequado de predição de híbridos simples de linhagens é o BLUP (Bernardo, 1999; Smith, 1999). Conforme relatado por Hallauer (1999), atualmente os híbridos simples são os tipos predominantes de híbridos desenvolvidos em milho e os métodos (média dos quatro híbridos simples não genitores) de predição usados para híbridos duplos (Jenkins, 1934) não são apropriados para híbridos simples, sendo recomendado, então, o procedimento BLUP. Também Smith (1999) refere-se à relevância do BLUP para o melhoramento das plantas anuais, revelando que os programas comerciais de melhoramento são essencialmente baseados em genealogia e, portanto, as informações de parentes são muito úteis e devem ser usadas na predição do mérito genotípico. Por este motivo, é seguro prever que o procedimento BLUP desempenhará um papel fundamental no melhoramento das plantas anuais, como atualmente desempenha no melhoramento de animais e de plantas perenes.

Os procedimentos apresentados a seguir são adequados à predição de híbridos de linhagens tanto de plantas perenes quanto anuais e já têm sido utilizados com sucesso no melhoramento do milho (Bernardo, 1996; 1999) e da soja (Panter e Allen, 1995).

O procedimento baseia-se no uso de médias por ensaio, de alguns híbridos avaliados em diferentes locais e anos. O BLUP corrige os dados para os efeitos fixos de ensaios (locais e anos) e, posteriormente, usa os dados corrigidos para a predição de híbridos não testados, através do uso das informações (parentesco) associadas à genealogia das linhagens genitoras e das variâncias

genéticas e residuais obtidas pelo procedimento da máxima verossimilhança restrita (REML).

No caso, tem-se o seguinte modelo linear (Resende et al., 1996a):

$$\bar{Y}_{ijk..} = \mu + \ell_k + H_{ij} + \frac{e_{ijk.}}{b} + \frac{\delta_{ijk..}}{nb}, \text{ em que:}$$

$\bar{Y}_{ijk..}$: produção média do híbrido entre as linhagens i e j, no ensaio k.

ℓ_k : efeito fixo do ensaio k.

H_{ij} : efeito aleatório do híbrido entre as linhagens i e j.

$\frac{e_{ijk.}}{b}$: efeito aleatório de parcela (efeito médio em relação ao número b de blocos).

$\frac{\delta_{ijk..}}{nb}$: efeito aleatório associado a indivíduo dentro de parcela (efeito médio em relação a nb, em que n é o número de plantas por parcela).

Considerando que $H_{ij} = G_i + G_j + D_{ij}$ e $e^* = \frac{e_{ijk.}}{b} + \frac{\delta_{ijk..}}{nb}$, o seguinte modelo pode ser reescrito, em termos matriciais.

$$y = X\beta + Zh + e^*$$

$$= X\beta + Z_1g_1 + Z_2g_2 + Zd_{12} + e^*, \text{ em que:}$$

y : vetor das médias por ensaio, dos híbridos entre linhagens dos grupos heteróticos 1 e 2.

β : vetor dos efeitos fixos de ensaio (locais e anos).

h: vetor dos efeitos dos híbridos, em que $Zh = Z_1g_1 + Z_2g_2 + Zd_{12}$.

g_1 : vetor dos efeitos genéticos da capacidade geral de combinação (metade dos efeitos aditivos interpопulacionais) de cada linhagem do grupo heterótico 1 em cruzamento com linhagens do grupo heterótico 2.

g_2 : vetor dos efeitos genéticos da capacidade geral de combinação (metade dos efeitos aditivos interpопulacionais) de cada linhagem do grupo heterótico 2 em cruzamento com linhagens do grupo heterótico 1.

d_{12} : vetor dos efeitos genéticos da capacidade específica de combinação entre linhagens dos grupos heteróticos 1 e 2.

e^* : vetor de efeitos residuais.

X, Z_1 , Z_2 e Z : matrizes de incidência (formadas por 1 e 0) para β , g_1 , g_2 e d_{12} , respectivamente.

É importante destacar que este modelo pode ser estruturado, também, em termos de médias por parcela (e não por ensaio) conforme especificado no tópico anterior.

Considerando a avaliação de n híbridos simples em ℓ ensaios provenientes de n_1 linhagens do grupo 1 e n_2 linhagens do grupo 2, num total de N observações, dimensões para os vetores são:

$$\beta = (\ell xl); \quad g_1 = (n_1 xl); \quad g_2 = (n_2 xl); \quad d_{12} = (nxl); \quad e = (Nx1); \quad y = (Nx1).$$

Se testemunhas forem incluídas nos ensaios, deve-se acrescentar no modelo Z_t t , em que t refere-se aos efeitos de testemunha e Z_t é a matriz de incidência para estes efeitos.

Neste caso, os preditores BLUP podem ser expressos em termos das equações de modelo misto de Henderson, por:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{d}_{12} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 & X'Z \\ Z_1'X & Z_1'Z_1 + A_1^{-1}\lambda_1 & Z_1'Z_2 & Z_1'Z \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1}\lambda_2 & Z_2'Z \\ Z'X & Z'Z_1 & Z'Z_2 & Z'Z + D^{-1}\lambda_3 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \\ Z'y \end{bmatrix}, \text{ em que:}$$

A_1 : matriz de parentesco (de Malecot) genético aditivo entre as linhagens do grupo 1 ($f_{AA'}$).

A_2 : matriz de parentesco (de Malecot) genético aditivo entre as linhagens do grupo 2 ($f_{BB'}$).

D : matriz de parentesco (de Malecot) genético de dominância entre as linhagens do grupo 1 e 2 ($f_{AA'} f_{BB'}$).

$$\lambda_1 = \frac{\sigma_{e^*}^2}{\sigma_{g_1}^2}; \quad \lambda_2 = \frac{\sigma_{e^*}^2}{\sigma_{g_2}^2}; \quad \lambda_3 = \frac{\sigma_{e^*}^2}{\sigma_{d_{12}}^2};$$

$\sigma_{e^*}^2 = \sigma_e^2/b + \sigma_\delta^2/nb$: variância residual.

$\sigma_{g_1}^2$: variância da capacidade geral de combinação interpopulacional que surge a partir do cruzamento das linhagens do grupo 1 com as do grupo 2.

$\sigma_{g_2}^2$: variância da capacidade geral de combinação interpopulacional que surge a partir do cruzamento das linhagens do grupo 2 com as do grupo 1.

$\sigma_{d_{12}}^2$: variância da capacidade específica de combinação interpopulacional.

Estes componentes de variância apresentam as seguintes relações com as variâncias genéticas aditivas e de dominância interpopulacionais:

$$\sigma_{g_1}^2 = [(1+F)/4]\sigma_{a_{12}}^2; \quad \sigma_{g_2}^2 = [(1+F)/4]\sigma_{a_{21}}^2; \quad \sigma_{d_{12}}^2 = [(1+F)/2]^2\sigma_{d_{(12)}}^2$$

Com linhagens completamente endogâmicas ($F=1$), têm-se as seguintes relações:

$\sigma_{a_{12}}^2 = 2 \sigma_{g_1}^2$: variância aditiva interpopulacional que surge a partir do cruzamento das linhagens do grupo 1 com as do grupo 2.

$\sigma_{a_{21}}^2 = 2 \sigma_{g_2}^2$: variância aditiva interpopulacional que surge a partir do cruzamento das linhagens do grupo 2 com as do grupo 1.

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

Estes componentes de variância podem ser estimados pelo método iterativo da máxima verossimilhança restrita –REML–, por meio do algoritmo EM via os seguintes estimadores:

$$\hat{\sigma}_{e^*}^2 = [y' y - \hat{\beta}' X' y - \hat{g}_1' Z_1' y - \hat{g}_2' Z_2' y - \hat{d}_{12}' Z' y] / [N - r(X)]$$

$$\hat{\sigma}_{g_1}^2 = [\hat{g}_1' A_1^{-1} \hat{g}_1 + \hat{\sigma}_{e^*}^2 \text{tr}(A_1^{-1} C^{11})] / n_1$$

$$\hat{\sigma}_{g_2}^2 = [\hat{g}_2' A_2^{-1} \hat{g}_2 + \hat{\sigma}_{e^*}^2 \text{tr}(A_2^{-1} C^{22})] / n_2$$

$$\hat{\sigma}_{d_{12}}^2 = [\hat{d}_{12}' D^{-1} \hat{d}_{12} + \hat{\sigma}_{e^*}^2 \text{tr}(D^{-1} C^{33})] / n, \text{ em que:}$$

tr : operador traço matricial.

r(X) : posto da matriz X.

C^{11} , C^{22} e C^{33} : representam as partições da inversa da matriz dos coeficientes referentes aos efeitos g_i , g_j e d_{ij} , respectivamente.

A estrutura de covariância genética apresentada advém do fato de que a covariância genética entre os híbridos $A \times B$ e $A' \times B'$, sendo A e A' linhagens do grupo 1 e B e B' linhagens do grupo 2, é dada por (Bernardo, 1996):

$$\text{COV}[(Ax B)(A' x B')] = f_{AA'} \hat{\sigma}_{g_1}^2 + f_{BB'} \hat{\sigma}_{g_2}^2 + f_{AA'} f_{BB'} \sigma_{d_{12}}^2$$

Os preditores apresentados fornecem os valores genotípicos preditos para os híbridos avaliados, pela expressão $\hat{h}_{ij} = \hat{g}_i + \hat{g}_j + \hat{d}_{ij}$. Fornecem também as estimativas REML dos componentes de variância aditiva e de dominância e as estimativas BLUE para os efeitos de ensaio ($\hat{\beta}$). De posse desses resultados, os valores genotípicos dos híbridos não testados são dados por: $h_{NT} = CV^{-1}yc$, em que:

$yc = (Z' Z)^{-1} Z' (y - X\hat{\beta})$: vetor de produções corrigidas (para os efeitos de ensaio) dos híbridos preditores.

C : matriz de covariância genética entre os híbridos não testados e os híbridos preditores (testados).

V : matriz de variância-covariância fenotípica entre os híbridos preditores ou testados.

Os elementos de C e os elementos não diagonais de V são dados pela expressão de $\text{COV}[(A \times B), (A' \times B')]$, refletindo, portanto, apenas componentes genéticos. Os elementos diagonais de V são componentes fenotípicos dados

por $\sigma_{g_1}^2 + \hat{\sigma}_{g_2}^2 + \sigma_{d_{12}}^2 + \sigma_{\epsilon^*}^2 / \ell_{ij}$, em ℓ_{ij} refere-se ao número de ensaios em que foi avaliado o híbrido ij.

Dessa forma, os valores genotípicos dos híbridos não testados são preditos a partir dos valores fenotípicos corrigidos dos híbridos avaliados.

Como exemplo, considere que cinco híbridos de café-arábica provenientes do cruzamento de duas linhagens do grupo heterótico um (x) e três linhagens do grupo heterótico dois (y) foram avaliados produzindo os seguintes resultados para o caráter peso de café-cereja, após avaliação em dois ensaios:

Ensaio	Híbrido	Produção de Café (kg/planta)
1	X ₁ - Y ₁ (1)	3,63
1	X ₁ - Y ₂ (2)	3,21
1	X ₁ - Y ₃ (3)	2,90
1	X ₂ - Y ₃ (4)	4,00
2	X ₁ - Y ₁ (1)	3,50
2	X ₂ - Y ₂ (5)	2,50

Neste caso, as seguintes matrizes de incidência e vetores são:

$$y = [3,63 \ 3,21 \ 2,90 \ 4,00 \ 3,50 \ 2,50]'$$

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

$$\beta = [\ell_1 \ \ell_2]'$$

$$Z_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix};$$

$g_1 = [g_{11} \ g_{12}]'$: contemplando as linhagens 1 do grupo 1 (g_{11}) e 2 do grupo 1 (g_{12}).

$$Z_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}; \quad g_2 = [g_{21} \ g_{22} \ g_{23}]'$$

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

$d_{12} = [d_1 \ d_2 \ d_3 \ d_4 \ d_5]'$: vetores dos efeitos de dominância dos 5 híbridos.

Considerando os componentes de variância

$\sigma_{e^*}^2 = 0,71$, $\sigma_{g_1}^2 = 0,94$, $\sigma_{g_2}^2 = 0,47$ e $\sigma_{d_{12}}^2 = 0,23$, tem-se:

$$\lambda_1 = \frac{0,71}{0,94} = 0,7553, \quad \lambda_2 = \frac{0,71}{0,47} = 1,5106, \quad \lambda_3 = \frac{0,71}{0,23} = 3,0870$$

Para o grupo 1 (X), na composição de A₁, pode-se considerar o parentesco de Malecot ou de coancestria: f₁₂ = 0,125.

Para o grupo 2 (Y), na composição de A₂, podem ser considerados os parentescos de Malecot ou de coancestria: f₁₂ = 0,08; f₁₃ = 0,06; f₂₃ = 0,00.

Para a composição dos elementos não diagonais de D, os parentescos de dominância são:

$$COV_d[(X_1 - Y_1), (X_1 - Y_2)] = f_{X1X1} f_{Y1Y2} = 1 \cdot 0,08 = 0,08$$

$$COV_d[(X_1 - Y_1), (X_1 - Y_3)] = f_{X1X1} f_{Y1Y3} = 1 \cdot 0,06 = 0,06$$

$$COV_d[(X_1 - Y_1), (X_2 - Y_3)] = f_{X1X2} f_{Y1Y3} = 0,125 \cdot 0,06 = 0,0075$$

$$COV_d[(X_1 - Y_1), (X_2 - Y_2)] = f_{X1X2} f_{Y1Y2} = 0,125 \cdot 0,08 = 0,01$$

$$COV_d[(X_1 - Y_2), (X_1 - Y_3)] = f_{X1X1} f_{Y2Y3} = 1 \cdot 0,00 = 0,00$$

$$COV_d[(X_1 - Y_2), (X_2 - Y_3)] = f_{X1X2} f_{Y2Y3} = 0,125 \cdot 0,00 = 0,00$$

$$COV_d[(X_1 - Y_2), (X_2 - Y_2)] = f_{X1X2} f_{Y2Y2} = 0,125 \cdot 1 = 0,125$$

$$COV_d[(X_1 - Y_3), (X_2 - Y_3)] = f_{X1X2} f_{Y3Y3} = 0,125 \cdot 1 = 0,125$$

$$COV_d[(X_1 - Y_3), (X_2 - Y_2)] = f_{X1X2} f_{Y3Y2} = 0,125 \cdot 0,0 = 0,00$$

$$COV_d[(X_2 - Y_3), (X_2 - Y_2)] = f_{X2X2} f_{Y3Y2} = 1 \cdot 0,00 = 0,00$$

Assim, as matrizes de parentesco de Malecot são:

$$A_1 = \begin{bmatrix} 1 & 0,125 \\ 0,125 & 1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1 & 0,08 & 0,06 \\ 0,08 & 1 & 0,00 \\ 0,06 & 0,00 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0,08 & 0,06 & 0,0075 & 0,01 \\ 0,08 & 1 & 0,00 & 0,00 & 0,125 \\ 0,06 & 0,00 & 1 & 0,125 & 0,00 \\ 0,0075 & 0,00 & 0,125 & 1 & 0,00 \\ 0,01 & 0,125 & 0,00 & 0,00 & 1 \end{bmatrix}$$

Efetuando-se as operações algébricas, obtém-se:

$$\begin{bmatrix} \hat{\beta} \\ g_1 \\ g_2 \\ d_{12} \end{bmatrix} = \begin{bmatrix} 1.3895 & 1.0004 & -0.8479 & -0.6416 & -0.1704 & -0.2257 & -0.3206 & -0.0751 & -0.0841 & -0.0668 & -0.1093 & -0.0402 & 1374 & 34448 \\ 1.0004 & 1.6834 & -0.7148 & -0.7746 & -0.2945 & -0.3020 & -0.1321 & -0.1324 & -0.0179 & -0.0455 & -0.0262 & -0.1484 & 600 & 30021 \\ -0.8479 & -0.7148 & 0.9499 & 0.5396 & -0.0704 & 0.0158 & 0.0490 & -0.0411 & -0.0273 & -0.0325 & 0.0598 & 0.0360 & 1324 & -0.0748 \\ -0.6416 & -0.7746 & 0.5396 & 0.9499 & 0.0704 & -0.0158 & -0.0490 & 0.0411 & 0.0273 & 0.0325 & -0.0598 & -0.0360 & 650 & 00748 \\ -0.1704 & -0.2945 & -0.0704 & 0.0704 & 0.5067 & 0.1364 & 0.1002 & -0.0774 & 0.0226 & 0.0306 & 0.0033 & 0.0233 & 713 & 01804 \\ -0.2257 & -0.3020 & 0.0158 & 0.1364 & 0.4687 & 0.1133 & 0.0414 & -0.0639 & 0.0286 & 0.0334 & -0.0425 & 571 & -0.1756 \\ -0.3206 & -0.1321 & 0.0490 & -0.0490 & 0.1002 & 0.1133 & 0.04953 & 0.0303 & 0.0420 & -0.0561 & -0.0359 & 0.0204 & 690 & 00071 \\ -0.0751 & -0.1324 & -0.0411 & 0.0411 & -0.0774 & 0.0114 & 0.0303 & 0.2853 & 0.0365 & 0.0339 & 0.0025 & 0.0155 & 713 & 00880 \\ -0.0841 & -0.0179 & -0.0273 & 0.0273 & 0.0226 & -0.0639 & 0.0420 & 0.0365 & 0.2887 & 0.0178 & 0.0052 & 0.0405 & 321 & 00003 \\ -0.0668 & -0.0455 & -0.0325 & 0.0325 & 0.0306 & 0.0286 & -0.0561 & 0.0339 & 0.0178 & 0.2847 & 0.0487 & -0.0021 & 290 & -0.01011 \\ -0.1093 & -0.0262 & 0.0598 & -0.0598 & 0.0033 & 0.0334 & -0.0359 & 0.0025 & 0.0052 & 0.0487 & 0.2959 & 0.0132 & 400 & 01043 \\ -0.0402 & -0.1484 & 0.0360 & -0.0360 & 0.0233 & -0.0425 & 0.0204 & 0.0155 & 0.0405 & -0.0021 & 0.0132 & 0.3005 & 250 & -0.0470 \end{bmatrix}$$

Para a predição do valor genotípico do híbrido (6) não testado ($X_2 - Y_1$), os parentescos de dominância são:

$$COV_d[(X_2 - Y_1), (X_1 - Y_1)] = f_{X_2 X_1} f_{Y_1 Y_1} = 0,125 \cdot 1 = 0,125$$

$$COV_d[(X_2 - Y_1), (X_1 - Y_2)] = f_{X_2 X_1} f_{Y_1 Y_2} = 0,125 \cdot 0,08 = 0,01$$

$$COV_d[(X_2 - Y_1), (X_1 - Y_3)] = f_{X_2 X_1} f_{Y_1 Y_3} = 0,125 \cdot 0,06 = 0,0075$$

$$COV_d[(X_2 - Y_1), (X_2 - Y_3)] = f_{X_2 X_2} f_{Y_1 Y_3} = 1 \cdot 0,06 = 0,06$$

$$COV_d[(X_2 - Y_1), (X_2 - Y_2)] = f_{X_2 X_2} f_{Y_1 Y_2} = 1 \cdot 0,08 = 0,08$$

As covariâncias genéticas entre o híbrido (6) ou ($X_2 - Y_1$) e os demais híbridos (1, 2,...,5), equivalem a:

$$\begin{aligned} COV[(6), (1)] &= COV_g[(X_2 - Y_1), (X_1 - Y_1)] = f_{X_2 X_1} \sigma_{g_1}^2 + f_{Y_1 Y_1} \sigma_{g_2}^2 + 0,125 \sigma_{d_{12}}^2 \\ &= 0,125 \sigma_{g_1}^2 + \sigma_{g_2}^2 + 0,125 \sigma_{d_{12}}^2 \\ &= 0,125 \cdot 0,94 + 0,47 + 0,125 \cdot 0,23 \\ &= 0,61625 \end{aligned}$$

$$COV[(6), (2)] = COV_g[(X_2 - Y_1), (X_1 - Y_2)] = f_{X_2 X_1} \sigma_{g_1}^2 + f_{Y_1 Y_2} \sigma_{g_2}^2 + 0,01 \sigma_{d_{12}}^2$$

$$COV[(6), (2)] = 0,125 \sigma_{g_1}^2 + 0,08 \sigma_{g_2}^2 + 0,01 \sigma_{d_{12}}^2 = 0,1574$$

$$\begin{aligned} COV[(6), (3)] &= COV_g[(X_2 - Y_1), (X_1 - Y_3)] = f_{X_2 X_1} \sigma_{g_1}^2 + f_{Y_1 Y_3} \sigma_{g_2}^2 + 0,0075 \sigma_{d_{12}}^2 \\ &= 0,125 \sigma_{g_1}^2 + 0,06 \sigma_{g_2}^2 + 0,0075 \sigma_{d_{12}}^2 = 0,1474 \end{aligned}$$

$$\begin{aligned} COV[(6), (4)] &= COV_g[(X_2 - Y_1), (X_2 - Y_3)] = f_{X_2 X_2} \sigma_{g_1}^2 + f_{Y_1 Y_3} \sigma_{g_2}^2 + 0,06 \sigma_{d_{12}}^2 \\ &= \sigma_{g_1}^2 + 0,06 \sigma_{g_2}^2 + 0,06 \sigma_{d_{12}}^2 = 0,982 \end{aligned}$$

$$COV(6,5) = COV_g[(X_2 - Y_1), (X_2 - Y_2)] = f_{X_2 X_2} \sigma_{g_1}^2 + f_{Y_1 Y_2} \sigma_{g_2}^2 + 0,08 \sigma_{d_{12}}^2$$

$$= \sigma_{g_1}^2 + 0,08 \sigma_{g_2}^2 + 0,08 \sigma_{d_{12}}^2 = 0,996$$

Assim:

$$C' = [0,6125 \ 0,1574 \ 0,1474 \ 0,982 \ 0,996]$$

$$V = \begin{bmatrix} 1,995 & 0,996 & 0,982 & 0,147425 & 0,1574 \\ 0,996 & 2,35 & 0,94 & 0,1175 & 0,61625 \\ 0,982 & 0,94 & 2,35 & 0,61625 & 0,1175 \\ 0,147425 & 0,1175 & 0,61625 & 2,35 & 0,94 \\ 0,1574 & 0,61625 & 0,1175 & 0,94 & 2,35 \end{bmatrix}$$

Os elementos diagonais de V foram obtidos por:

$V_{11} = \sigma_{g_1}^2 + \sigma_{g_2}^2 + \sigma_{d_{12}}^2 + \sigma_{e^*}^2 / 2 = 1,995$, visto que o híbrido (1) foi avaliado em 2 ensaios;

$V_{ii} = \sigma_{g_1}^2 + \sigma_{g_2}^2 + \sigma_{d_{12}}^2 + \sigma_{e^*}^2 = 2,35$ para os demais ($i > 1$) elementos, já que os demais híbridos foram avaliados em apenas um ensaio.

Os elementos não diagonais de V foram assim obtidos:

$$COV(1,2) = \sigma_{g_1}^2 + 0,08 \sigma_{g_2}^2 + 0,08 \sigma_{d_{12}}^2 = 0,996$$

$$COV(1,3) = \sigma_{g_1}^2 + 0,06 \sigma_{g_2}^2 + 0,06 \sigma_{d_{12}}^2 = 0,982$$

$$COV(1,4) = 0,125 \sigma_{g_1}^2 + 0,06 \sigma_{g_2}^2 + 0,0075 \sigma_{d_{12}}^2 = 0,147425$$

$$COV(1,5) = 0,125 \sigma_{g_1}^2 + 0,08 \sigma_{g_2}^2 + 0,01 \sigma_{d_{12}}^2 = 0,1574$$

$$COV(2,3) = \sigma_{g_1}^2 + 0,0 \sigma_{g_2}^2 + 0,00 \sigma_{d_{12}}^2 = 0,94$$

$$COV(2,4) = 0,125 \sigma_{g_1}^2 + 0,0 \sigma_{g_2}^2 + 0,00 \sigma_{d_{12}}^2 = 0,1175$$

$$COV(2,5) = 0,125 \sigma_{g_1}^2 + \sigma_{g_2}^2 + 0,125 \sigma_{d_{12}}^2 = 0,61625$$

$$COV(3,4) = 0,125 \sigma_{g_1}^2 + \sigma_{g_2}^2 + 0,125 \sigma_{d_{12}}^2 = 0,61625$$

$$COV(3,5) = 0,125 \sigma_{g_1}^2 + 0,0 \sigma_{g_2}^2 + 0,00 \sigma_{d_{12}}^2 = 0,1175$$

$$COV(4,5) = \sigma_{g_1}^2 + 0,0 \sigma_{g_2}^2 + 0,0 \sigma_{d_{12}}^2 = 0,94$$

Posteriormente, podem ser obtidos:

$$yc = (Z'Z)^{-1}Z'(y - X\hat{\beta}) = \begin{bmatrix} 0,3416 \\ -0,2348 \\ -0,5448 \\ 0,5552 \\ -0,5021 \end{bmatrix}$$

$$h_{NT} = CV^{-1}yc = 0,2635$$

Assim, o efeito genotípico do híbrido (6) é 0,2635 kg. O valor genotípico dos híbridos é dado pela soma dos efeitos genotípicos com a média geral (3,22345). Dessa forma, o valor genotípico do híbrido (6) equivale a 3,22345 + 0,2635 = 3,48695 kg, e o do híbrido (1) é :

$$\mu + \hat{g}_{11} + \hat{g}_{21} + \hat{d}_1 = 3,22345 - 0,0748 + 0,1804 + 0,088 = 3,4171 \text{ kg}$$

Os valores -0,0741; 0,1804 e 0,088 foram obtidos do vetor de solução.

$$\begin{bmatrix} \hat{\beta} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{d}_{12} \end{bmatrix} = \begin{bmatrix} 3,4448 \\ 3,0021 \\ -0,0748 \\ 0,0748 \\ 0,1804 \\ -0,1756 \\ 0,0071 \\ 0,0880 \\ 0,0003 \\ -0,1011 \\ 0,1043 \\ -0,0470 \end{bmatrix}, \text{ calculado anteriormente. O valor } 3,22345 \text{ equivale à média dos 2 primeiros elementos deste vetor.}$$

É interessante observar que, embora apenas recentemente o procedimento BLUP começou a ser aplicado ao melhoramento de plantas anuais na seleção de cruzamentos entre linhagens, este foi o primeiro tema a ser abordado por Henderson em conexão com a metodologia de modelos mistos. Tal foi o tema de sua tese de doutorado (Henderson, 1948) e de outras de suas várias pesquisas (Henderson, 1952; 1977, 1984). Os procedimentos aqui relatados podem ser estendidos para a predição de híbridos não observados em espécies alógamas perenes. Tais híbridos, em geral, advêm de genitores não endogâmicos.

17.4. Uso de Progênies de Autofecundação em Espécies Alógamas

A seleção baseada em progênies de polinização aberta (meios-irmãos para espécies alógamas) tem sido praticada de forma intensiva no melhoramento de plantas perenes. Isto se deve à facilidade de obtenção deste tipo de progénie, à possibilidade de estimação da herdabilidade no sentido restrito e ao grande potencial genético dos métodos de seleção baseados em testes deste tipo de progénie, visto que a utilização das fontes de informação entre e dentro de progênies possibilita explorar toda a variação genética aditiva das populações.

A utilização de outros tipos de progênies que permitem explorar maior quantidade de variação genética tem despertado a atenção dos melhoristas. Dentro destes tipos, as progênies de autofecundação (S_1) apresentam potencial, visto que uma geração de autofecundação aumenta a variação genética aditiva da população em 50% ($3/2 \sigma_a^2$). Entretanto, a endogamia em espécies alógamas conduz ao aparecimento de três (D_1 , D_2 e H) novos componentes de variação genética, além da variação aditiva e de dominância (Cockerham, 1983), os quais participam efetivamente da expressão do progresso genético esperado com seleção. Portanto, na comparação entre a efetividade da seleção baseada em progênies não endógamas e S_1 , deve-se considerar estes três novos componentes.

São apresentados neste tópico alguns resultados relativos à comparação da eficiência seletiva de progênies de meios-irmãos e S_1 's, conforme obtido por Resende & Vencovsky (1992; 1993). No estudo foram consideradas as seguintes expressões do ganho genético:

(a) Seleção entre progênies de meios-irmãos

$$G_s = \frac{k(1/4) \sigma_a^2}{[(1/4) \sigma_a^2 + \sigma_e^2/b + (3/4) \sigma_a^2 + \sigma_d^2 + \sigma_{ad}^2]/(nb)]^{1/2}}$$

(b) Seleção dentro de progênies de meios-irmãos

$$G_s = \frac{k(3/4) \sigma_a^2}{[(3/4) \sigma_a^2 + \sigma_d^2 + \sigma_{ad}^2]^{1/2}}$$

(c) Seleção entre progênies S_1

Herdabilidade da média de progénie

$$\begin{aligned} h_f^2 &= \frac{\sigma_a^2 + (1/2) D_1}{\sigma_{F^*}^2} = \\ &= \frac{\sigma_a^2 + (1/2) D_1}{[(\sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (1/8) D_2) + \sigma_e^2/b + (1/2) \sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (3/8) D_2 + (1/4) H + \sigma_{ad}^2]/(nb)]} \end{aligned}$$

Ganho genético

$$G_s = k h_f^2 \sigma_{F^*} = \frac{k[\sigma_a^2 + (1/2) D_1]}{\sigma_{F^*}}$$

(d) Seleção dentro de progênies S₁**Herdabilidade de indivíduo dentro de parcela**

$$h_{\delta}^2 = \frac{(1/2) \sigma_a^2 + (1/2) D_1}{\sigma_{\delta^*}^2} = \frac{(1/2) \sigma_a^2 + (1/2) D_1}{(1/2) \sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (3/8) D_2 + (1/4) H + \sigma_{ad}^2}$$

Ganho genético

$$G_s = kh_{\delta}^2 \sigma_{\delta^*} = \frac{K[(1/2) \sigma_a^2 + (1/2) D_1]}{\sigma_{\delta^*}}$$

(e) Seleção combinada baseada em progênies de meios-irmãos

$$G_s = k \left[\frac{\left[\frac{1+(nb-1)(1/4)}{nb} \sigma_a^2 \right]^2}{(1/4) \sigma_a^2 + \frac{(3/4) \sigma_a^2 + \sigma_d^2 + \sigma_{ad}^2}{nb} + \frac{\sigma_e^2}{b}} + \frac{\frac{n-1}{n} [(3/4) \sigma_a^2]^2}{(3/4) \sigma_a^2 + \sigma_d^2 + \sigma_{ad}^2} \right]^{1/2}$$

(f) Seleção combinada baseada em progênies S₁

$$G_s = k \left[\frac{\left[\frac{2nb+1}{2nb} \sigma_a^2 + \frac{nb-1}{2nb} D_1 \right]^2}{\sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (1/8) D_2 + \frac{[(1/2) \sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (3/8) D_2 + (1/4) H + \sigma_{ad}^2]}{nb} + \frac{\sigma_e^2}{b}} + \frac{\frac{n-1}{n} [(1/2) \sigma_a^2 + (1/2) D_1]^2}{(1/2) \sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (3/8) D_2 + (1/4) H + \sigma_{ad}^2} \right]^{1/2}$$

(g) Seleção massal em teste de progênies de meios-irmãos

$$G_s = \frac{k \sigma_a^2}{[\sigma_a^2 + \sigma_d^2 + \sigma_e^2 + \sigma_{ad}^2]^{1/2}}$$

(h) Seleção massal com progênies S₁

$$G_s = \frac{k [(3/2) \sigma_a^2 + D_1]}{[(3/2) \sigma_a^2 + (1/2) \sigma_d^2 + 2D_1 + (1/2) D_2 + (1/4) H + \sigma_e^2 + \sigma_{ad}^2]^{1/2}}$$

(i) Seleção de genitores com base em suas progênies de meios-irmãos

$$G_s = \frac{k (1/2) \sigma_a^2}{[(1/4) \sigma_a^2 + [(3/4) \sigma_a^2 + \sigma_d^2 + \sigma_{ad}^2]/(nb) + \sigma_e^2/r]^{1/2}}$$

(j) Seleção de genitores com base em suas progênies S₁

$$G_s = \frac{k [\sigma_a^2 + (1/2) D_1]}{[\sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (1/8) D_2 + [(1/2) \sigma_a^2 + (1/4) \sigma_d^2 + D_1 + (3/8) D_2 + (1/4) H + \sigma_{ad}^2]/(nb) + \sigma_e^2/b]^{1/2}}$$

(k) Seleção entre progênies de meios-irmãos obtidas de plantas S₁

$$G_s = \frac{k (3/8) \sigma_a^2}{[(3/8) \sigma_a^2 + [(5/8) \sigma_a^2 + \sigma_d^2 + \sigma_{ad}^2]/(nb) + \sigma_e^2/b]^{1/2}}$$

(I) Seleção dentro de progêneres de meios-irmãos obtidas de plantas S₁

$$G_s = \frac{k (5/8) \sigma_a^2}{[(5/8) \sigma_a^2 + \sigma_d^2 + \sigma_{ad}^2]^{1/2}}$$

Para a obtenção dos componentes de variância genética utilizou-se o mesmo procedimento adotado por Fernandes (1990) em que as diferentes freqüências alélicas (caráter quantitativo) de uma determinada população se ajustam a uma distribuição Beta. A partir das expressões dos componentes da variância genética foram obtidos os valores esperados dos parâmetros populacionais de três tipos de populações quanto ao grau de melhoramento, a saber: pouco melhorada, nível intermediário de melhoramento, e melhorada, com freqüências médias dos alelos fatoráveis em torno de 0,3; 0,5 e 0,7, respectivamente. Considerou-se um valor base de variância aditiva equivalente a 1,845, o qual foi tomado como referente a uma população com $\bar{p} = 0,5$ e grau médio de dominância 1. Os demais parâmetros foram obtidos a partir da relação esperada deles com esta variância aditiva base.

Considerando os graus médios de dominância 0,75 (dominância parcial) e 1,0 (dominância completa), os valores dos parâmetros são apresentados na Tabela 17.3.

Tabela 17.3. Valores dos componentes de variância genética utilizados para as seis populações (combinação de três freqüências alélicas médias e dois níveis de dominância), bem como herdabilidades individuais no sentido restrito (h_a^2) associadas, considerando $\sigma_{ad}^2 = 12,75$ e $\sigma_e^2 = 1,5$.

	\bar{p}	0,75	h_a^2	1,00	h_a^2
0,3	σ_a^2	2,611		2,844	
	σ_d^2	0,377		0,632	
	D ₁	-0,573	0,15	-0,790	0,16
	D ₂	0,329		0,553	
0,5	H	0,377		0,632	
	σ_a^2	1,845		1,845	
	σ_d^2	0,412		0,692	
	D ₁	-0,137	0,11	-0,230	0,11
0,7	D ₂	0,275		0,460	
	H	0,412		0,642	
	σ_a^2	0,981		0,789	
	σ_d^2	0,377		0,632	
0,7	D ₁	0,243	0,06	0,237	0,05
	D ₂	0,329		0,553	
	H	0,377		0,632	

Observa-se que a variância genética aditiva, em relação à população com nível intermediário de melhoramento, aumenta para populações pouco melhoradas e diminui para populações melhoradas, independentemente do nível de dominância. Isto evidencia o maior potencial de resposta à seleção de populações pouco melhoradas em relação às mais melhoradas.

A variância genética de dominância e o componente D₂ apresentam valores simétricos em relação à freqüência alélica média de 0,5, ou seja, seus valores são iguais para as freqüências 0,3 a 0,7. A variância de dominância apresenta maiores valores na freqüência alélica de 0,5, independentemente do nível de dominância. Entretanto, em uma mesma freqüência alélica, a variância genética de dominância aumenta com o grau médio de dominância (logicamente), a ponto de se aproximar da magnitude da variância genética aditiva na freqüência de 0,7 e grau médio de dominância de 1,0.

O componente H apresenta valor idêntico ao de σ_a^2 , pois foi considerado o modelo com dois alelos por locos, conforme Cockerham & Weir (1984).

Dos novos componentes que surgem com a endogamia, o que parece chamar mais a atenção é D₁, pois apresentou valores negativos nas freqüências alélicas de 0,3 e 0,5. Além do mais, é o único dos novos componentes que participa do numerador da expressão do ganho genético com progênies S₁. Assim, verifica-se que nesses casos o componente D₁ é subtraído da variância genética aditiva explorável pela seleção, de forma que o simples exame visual da expressão do ganho não permite concluir se o numerador da expressão de ganho com seleção baseada em S₁ é superior ao de meios-irmãos. Para esta comparação, deve-se considerar sobretudo as magnitudes relativas entre a variância genética aditiva e o componente D₁. Nesse aspecto, constata-se que a maior magnitude de D₁, em relação a σ_a^2 , ocorre na freqüência alélica de 0,3 e grau médio de dominância 1,0, situação esta que deverá ser a mais complicadora para S₁, em comparação com meios-irmãos.

Considerando nb = 30 plantas por progénie e precisão experimental idêntica para meios-irmãos e S₁, os resultados comparativos são apresentados nas Tabelas 17.4 e 17.5, para os métodos de seleção entre progênies, dentro de progênies e entre e dentro de progênies.

Tabela 17.4. Eficiências relativas (ER) da seleção entre progênies, dentro de progênies e entre e dentro de progênies baseadas em S₁ e meios-irmãos para caracteres com grau de dominância igual a 0,75 em três populações com vários graus de melhoramento e variadas intensidades (i) de seleção.

Freqüência alélica média = 0,3

	Entre	Dentro	Entre e Dentro
i	0,10	0,10	0,01
ER	2,51	0,55	1,51
i	0,10	0,02	0,002
ER	2,51	0,55	1,35

Freqüência alélica média = 0,5

	Entre	Dentro	Entre e Dentro
i	0,10	0,10	0,01
ER	2,68	0,63	1,66
i	0,10	0,02	0,002
ER	2,68	0,63	1,50

Freqüência alélica média = 0,7

	Entre	Dentro	Entre e Dentro
i	0,10	0,10	0,01
ER	3,11	0,83	2,02
i	0,10	0,02	0,002
ER	3,11	0,83	1,83

Tabela 17.5. Eficiências relativas (ER) da seleção entre progênies, dentro de progênies e entre e dentro de progênies baseadas em S_1 e meios-irmãos para caracteres com grau médio de dominância igual a 1,0 em três populações com variados graus de melhoramento e variadas intensidades (i) de seleção.

Freqüência alélica média = 0,3

	Entre	Dentro	Entre e Dentro
i	0,10	0,10	0,01
ER	2,41	0,51	1,44
i	0,10	0,02	0,002
ER	2,41	0,51	1,29

Freqüência alélica média = 0,5

	Entre	Dentro	Entre e Dentro
i	0,10	0,10	0,01
ER	2,59	0,60	1,60
i	0,10	0,02	0,002
ER	2,59	0,60	1,44

Freqüência alélica média = 0,7

	Entre	Dentro	Entre e Dentro
i	0,10	0,10	0,01
ER	3,14	0,87	2,07
i	0,10	0,02	0,002
ER	3,14	0,87	1,87

Verifica-se pelas Tabelas 17.4 e 17.5, que em todas as situações o progresso genético pela seleção entre progênies foi sempre superior para S₁, enquanto que o progresso pela seleção dentro de progênies foi sempre superior para meios-irmãos. O ganho genético total pela seleção entre e dentro de progênies foi superior para S₁ em todas as situações, sendo mais vantajosa para este tipo de progénie, quanto mais melhoradas forem as populações. Isto ocorre devido ao fato de o componente D₁ apresentar sinal negativo nas freqüências alélicas intermediárias e baixas.

Para populações pouco melhoradas e com níveis intermediários de melhoramento, constatou-se que as menores eficiências foram verificadas para níveis mais altos de dominância alélicas, situações essas em que o componente D₁ apresentou maiores magnitudes em relação à variância aditiva. Já nas populações mais melhoradas houve menores influências do grau médio de dominância, pois nestas situações o componente D₁ apresentou sinais positivos.

Analizando-se a eficiência relativa em função das combinações de intensidade de seleção entre e dentro percebe-se que a melhor situação para S₁'s foi aquela em que manteve-se a intensidade de seleção dentro (10%) na mesma taxa de seleção entre (10%). Isto confirma que o grande mérito da seleção com S₁'s é realmente na seleção entre progênies. Porém, mesmo com uma alta intensidade de seleção dentro (2%), a seleção com S₁'s mostrou-se bastante superior.

Os resultados de eficiência apresentados são, provavelmente, superestimados, uma vez que a precisão experimental para S₁ deve ser mais baixa, visto que progênies endógamas podem ser mais susceptíveis a variações ambientais do que as não endógamas (Vencovsky, 1987a). Em função disto, considerou-se também a variação ambiental para S₁ como sendo o dobro dos valores para meios-irmãos, e os resultados obtidos encontram-se na Tabela 17.6.

Tabela 17.6. Eficiências relativas (ER) da seleção entre e dentro de progênies S₁ sobre a seleção entre e dentro de progênies de meios-irmãos (ER₁) e seleção combinada com S₁ sobre combinada com meios-irmãos (ER₂), para as seis populações (três freqüências alélicas – FA e dois graus médios de dominância – GMD) estudadas, considerando a variância ambiental para S₁ como sendo o dobro daquela para meios-irmãos e intensidades de seleção de 10% entre e 2% dentro de progênies.

FA	GMD	h_a^2	ER ₁	ER ₂	ER _{3*}	ER**
0,3	0,75	0,15	1,15	-	-	1,12
0,5	0,75	0,11	1,26	-	-	1,18
0,7	0,75	0,06	1,51	-	-	1,34
0,3	1,00	0,16	1,10	1,24	1,08	1,08
0,5	1,00	0,11	1,22	1,47	1,17	1,14
0,7	1,00	0,05	1,54	1,93	1,45	1,34

* Seleção massal; **: seleção de genitores.

Verifica-se pela Tabela 17.6, que a seleção com progêneres S_1 ainda continua sendo vantajosa sobre a de meios-irmãos, mas a taxa bem reduzidas, exceto para populações melhoradas. Dentre os vários métodos, o que apresenta menor eficiência para S_1 é a seleção de genitores.

Em algumas populações, em comparação com a seleção entre e dentro, a seleção combinada, baseada em S_1 's é ainda mais vantajosa em relação à de meios-irmãos. Isto ocorre em função de a seleção combinada enfatizar mais a herdabilidade no que se refere às médias (que é maior que a dentro de famílias) do que o faz a seleção entre e dentro (que no presente caso advém de uma intensidade de seleção de 10% entre e 2% dentro). Como a herdabilidade em termos de médias de famílias é muito maior para S_1 's e a dentro de famílias é maior para meios-irmãos, a eficiência da seleção com S_1 em relação à de meios-irmãos aumenta na seleção combinada em relação àquelas entre e dentro de progêneres.

Vale ressaltar que diferentes métodos de melhoramento são melhor comparados quando se levam em consideração o tamanho efetivo e o incremento da endogamia que eles acarretam (Vencovsky (1978)). Entretanto, conforme Vencovsky (1987a), considerar tamanhos efetivos iguais ou ignorá-los na comparação de métodos de seleção é uma questão de estratégia. No presente trabalho, as intensidades de seleção foram iguais e, consequentemente, os tamanhos efetivos associados a cada modalidade de seleção (S_1 's e meios-irmãos), diferentes. No entanto, em espécies perenes, altos ganhos são desejados em cada ciclo (progresso genético imediato) em função deste ser muito longo, de forma que optou-se por realizar as comparações em termos de ganho genético por ciclo, independentemente de estarem associados a tamanhos efetivos diferentes. Logicamente, que um tamanho efetivo mínimo deve ser mantido para garantir a obtenção dos tetos seletivos das populações, após vários ciclos de seleção. Como o tamanho efetivo associado à seleção com S_1 's é menor que aquele associado à seleção com meios-irmãos, deve-se avaliar, inicialmente, maior número de famílias S_1 em relação ao número que seria avaliado de famílias de meios-irmãos. Isso permitiria manter o tamanho efetivo mínimo necessário, sem prejudicar a intensidade de seleção.

Outra questão que deve ser analisada também é o comprimento do ciclo da seleção associado a S_1 's e meios-irmãos. Analisando um único ciclo de seleção, o tempo necessário para gerar material melhorado é o mesmo em ambos os processos seletivos, pois a recombinação, na maioria dos casos, será realizada por meio dos indivíduos selecionados a uma determinada idade comum. Porém, analisando o ciclo subsequente, verifica-se que para meios-irmãos o processo de obtenção das progêneres coincide com a recombinação do ciclo anterior enquanto que para a obtenção de novas progêneres S_1 , faz-se necessária a recombinação prévia dos S_1 's selecionados no ciclo anterior, o que conduziria à obtenção de dois ciclos de seleção em três gerações, fato que poderia representar uma grande desvantagem em relação a meios-irmãos.

Para a herdabilidade de 11%, a eficiência da seleção combinada com S₁ foi de 1,47 (Tabela 17.6). Para herdabilidade na mesma magnitude, Resende (1999b) verificou que a eficiência de delineamento de cruzamento fatorial 4 x 4 sobre a seleção combinada com meios-irmãos foi de 1,39. Considerando os problemas (baixo tamanho efetivo, maior número de gerações por ciclo, baixa produção de sementes, ausência de coincidência de florescimento entre os vários indivíduos S₁ por ocasião da recombinação, possíveis problemas experimentais) associados à utilização de S₁, pode-se afirmar que a utilização do fatorial desconexo 4 x 4 é uma opção muito mais segura e que apresenta ganho genético similar ao ganho potencial com a utilização de S₁. Tal delineamento é adequado, também, à seleção de genitores.

É importante ressaltar que o modelo genético utilizado não leva em consideração a presença de alelos deletérios (carga genética) de forma que, na prática, as médias de determinadas progênies poderiam ser prejudicadas pelo aparecimento de determinados indivíduos atípicos, ou seja, determinado genitor poderia apresentar uma alta freqüência de alelos favoráveis e ser descartado por possuir um ou poucos alelos deletérios. Entretanto, conforme proposto por Barker & Libby (1974), a seleção entre progênies baseada apenas nos 50% indivíduos superiores de cada família poderia reduzir os efeitos da presença de genes maiores deletérios e, consequentemente, permitir uma classificação mais fidedigna dos melhores genitores sob alogamia. Neste contexto, a alternativa seria o uso de S₁'s apenas em populações já melhoradas (com menos alelos letais e deletérios).

Ainda com relação ao modelo genético utilizado, pode-se dizer que este é ainda simplificado, pois não leva em consideração os componentes de variância referentes à epistasia. A adoção deste modelo simplificado é satisfatório para o caso de completa alogamia. Quando a endogamia ocorre em espécies naturalmente alógamas, os componentes da variância genética epistática podem assumir relevância. Uma demonstração empírica disso pode ser vista no trabalho de Wilcox (1983), que em estudo comparativo de famílias de polinização aberta e autofecundadas de *Pinus radiata* concluiu que o modelo genético sem epistasia mostrou-se inadequado para explicar as estruturas de variâncias observadas.

Finalmente, é importante ressaltar o possível potencial do uso de S₁, porém é necessário adquirir experiência prática com este tipo de família em espécies perenes.

Uma alternativa para contornar a questão do maior ciclo seletivo com S₁'s é a utilização de processo seletivo intercalando a utilização de progênies S₁'s e progênies de meios-irmãos obtidas de plantas S₁'s o que conduz à obtenção de um ciclo de seleção a cada geração. É importante ressaltar que a seleção (entre e dentro) com meios-irmãos a ser utilizada é superior à seleção com meios-irmãos obtidos de plantas não endógamas, fornecendo eficiências relativas (ER) de 1,06; 1,07 e 1,10 para caracteres com graus de dominância 1,0 em populações pouco melhoradas, com nível intermediário de melhoramento e melhora-

das, respectivamente (intensidades de seleção de 10% entre e 2% dentro). Em suma, o processo seletivo combinando a utilização de dois tipos de progêneres que, isoladamente, dão ganho superior à tradicional seleção com meios-irmãos, sem dúvida, é desejável.

Para contornar a questão do tamanho efetivo, um procedimento baseado em dois tipos de progêneres poderia ser delineado, praticando-se a seleção entre progêneres com base em S_1 (ou em índice usando simultaneamente S_1 e meios-irmãos) avaliadas em experimentos e a seleção dentro de progêneres e recombinação com base em meios-irmãos, avaliados em parcelas fora do experimento.

É importante relatar, também, que a eficiência do uso de S_1 dependerá sobretudo da estrutura genética de cada particular população. Em milho, Lamkey (1999) relata que a seleção baseada em progêneres endogâmicas foi efetiva para melhorar algumas populações mas, em outras, foi ineficiente.

Capítulo 18

**Espécies com Sistema
Reprodutivo Misto**



18.1. Introdução

Espécies com sistema reprodutivo misto, que praticam simultaneamente autofecundações e cruzamentos, são comuns dentre as perenes. As mais importantes economicamente são a seringueira, o eucalipto, o cacauzeiro e o coqueiro (Baril & Bergonzini ,1995a). Espécies anuais como o algodão, o girassol, o sorgo, o melão, a berinjela e o quiabo também apresentam sistema reprodutivo misto.

Nestas espécies, a suposição de que a covariância (cov_p) entre indivíduos de uma progênie de polinização aberta equivale a 25% da variância genética aditiva (σ_a^2) tem sido um fator que, geralmente, ocasiona superestimativas dessa variância, bem como do coeficiente de herdabilidade e dos ganhos genéticos com seleção (Namkoong, 1966; Squillace, 1974). Isto ocorre devido ao fato de que as autofecundações aumentam a covariância entre os indivíduos da progênie, de forma que empregando-se a relação $\sigma_a^2 = 4 \text{ cov}_p$, provoca-se uma superestimação de σ_a^2 .

A suposição de que progênies de polinização aberta são compostas por meios-irmãos não é válida para espécies de sistema reprodutivo misto, portanto é necessária a adoção de técnicas que possam contornar o problema. Diversos autores têm proposto formas alternativas de considerar os efeitos decorrentes da autofecundação. Namkoong (1966) propõe apenas a correção da estimativa da variância genética aditiva com base no coeficiente de endogamia. Squillace (1974) apresenta multiplicadores da variância genética aditiva obtidos com base em covariâncias genéticas médias entre indivíduos de uma progênie de polinização aberta, os quais apresentam as relações de parentesco: meios-irmãos, irmãos germanos, irmãos S_1 , e irmãos cruzados e autofecundados.

Tais alternativas não consideram o modelo genético próprio para as espécies de sistema reprodutivo misto, modelo este apresentado por Cockerham & Weir (1984). Este modelo possibilita melhor caracterização da estrutura genética das populações e, indubitavelmente, propicia uma definição mais adequada dos componentes da variação genética e, assim, dos efeitos da autofecundação na estimação dos parâmetros genéticos. É importante ressaltar que as considerações realizadas por Namkoong (1966) e Squillace (1974) baseiam-se apenas na variância genética aditiva, não considerando os demais componentes da variação genética que surgem na presença de endogamia.

18.2. Modelo Genético

No modelo genético de Cockerham & Weir (1984), a variância genética total (σ_g^2) de uma população com sistema misto de reprodução equivale a $\sigma_g^2 = (1+F) \sigma_a^2 + (1-F) \sigma_d^2 + 4FD_1 + FD_2 + F(1-F)H + (F_c - F^2)(H^2 - H)$, em que F é o coeficiente de endogamia de Wright (obtido a partir da taxa de autofecundação S : F

= S/(2-S)) e F_c é o coeficiente de endogamia conjunto para dois locos, equivalendo a $F_c = [F(1+2F)]/(2+F)$. Definem-se os seguintes parâmetros:

σ_a^2 : variância genética aditiva.

σ_d^2 : variância genética de dominância.

D_1 : covariância entre os efeitos aditivos dos alelos e os efeitos de dominância dos homozigotos.

D_2 : variância genética dos efeitos de dominância dos homozigotos.

H : soma dos quadrados da depressão por endogamia.

H^2 : quadrado da soma da depressão por endogamia.

Neste modelo, a covariância genética entre parentes é expressa como função linear dos diversos componentes da variância genética. Assim, na obtenção das expressões dos progressos genéticos esperados com seleção, determina-se a covariância genética [$\text{cov}_g(x, y)$] entre a unidade de seleção (x) e um indivíduo da população melhorada (y), por meio da expressão apresentada por Cockerham & Weir (1984) e conforme notação utilizada por Pereira (1989):

$$\begin{aligned} \text{cov}_g(x, y) = & 2 f_{xy} \sigma_a^2 + 2 (T_{x+y} - t_{x'y'}) \sigma_d^2 + 2 (G_{x'y} + G_{xy'}) D_1 + t_{x'y'} D_2 + \\ & + (T_{x'y'} - F_x F_y) H + (T_{cxy} - F_x F_y) (H^2 - H) \end{aligned}$$

Considerando o indivíduo x com os alelos x_1 e x_2 e o indivíduo y com os alelos y_1 e y_2 , as medidas de identidade para um loco, relatadas acima, são definidas a seguir, em que pr denota a probabilidade de identidade por descendência indicada pelo sinal \equiv (Cockerham, 1971; Weir & Cockerham, 1977; Cockerham & Weir, 1984).

$$4 f_{xy} = \text{pr}(x_1 \equiv y_1) + \text{pr}(x_1 \equiv y_2) + \text{pr}(x_2 \equiv y_1) + \text{pr}(x_2 \equiv y_2)$$

$$2 T_{x+y} = \text{pr}(x_1 \equiv y_1 \text{ e } x_2 \equiv y_2) + \text{pr}(x_1 \equiv y_2 \text{ e } x_2 \equiv y_1)$$

$$t_{x'y'} = \text{pr}(x_1 \equiv x_2 \equiv y_1 \equiv y_2)$$

$$2 G_{x'y} = \text{pr}(x_1 \equiv x_2 \equiv y_1) + \text{pr}(x_1 \equiv x_2 \equiv y_2)$$

$$2 G_{xy'} = \text{pr}(y_1 \equiv y_2 \equiv x_1) + \text{pr}(y_1 \equiv y_2 \equiv x_2)$$

$$T_{x'y} = \text{pr}(x_1 \equiv x_2 \text{ e } y_1 \equiv y_2)$$

$$F_x = \text{pr}(x_1 \equiv x_2)$$

$$F_y = \text{pr}(y_1 \equiv y_2)$$

T_{cxy} = probabilidade de os dois indivíduos x e y , terem os alelos no loco i em um e os alelos no loco j em outro, idênticos por descendência.

As expressões apresentadas são válidas, na suposição de populações em equilíbrio e ausência de ligação gênica e epistasia.

De acordo com Wright & Cockerham (1985), populações de sistema reprodutivo misto não retornam ao equilíbrio após uma geração de cruzamentos ao acaso, mas sim, após várias gerações, em função da taxa de autofecundação.

ção. Dessa forma, conforme esses autores, têm-se dois tipos de ganho genético pela seleção, um denominado imediato (correspondente ao progresso na geração imediata após a seleção) e outro permanente (verificado quando a população selecionada atinge o equilíbrio).

18.3. Progresso Genético e Métodos de Seleção

Nas espécies perenes, pelos esquemas de produção de sementes normalmente utilizados, ou seja, logo após cada seleção são estabelecidos os pômares permanentes de produção de sementes, apenas o progresso genético imediato é aplicável, visto que uma população selecionada nunca é cruzada aleatoriamente por várias gerações.

A expressão do progresso genético (G_s) imediato com seleção equivale a (Wright & Cockerham, 1986; Baril & Bergonzini, 1995b):

$$G_s = [k \operatorname{cov}_g(x, y) / \sigma_x] + (\mu_x - \mu), \text{ em que:}$$

k : diferencial de seleção padronizado.

σ_x : desvio padrão fenotípico da unidade de seleção x .

μ : média da população em equilíbrio.

μ_x : média da população do experimento.

Segundo os autores, o primeiro termo da expressão reflete o efeito da seleção e o segundo refere-se à mudança do sistema de cruzamento utilizado para produzir as progênies. Assim, tratando-se progênies de polinização aberta, o segundo termo da expressão equivale a zero.

Objetivando verificar a influência do sistema reprodutivo misto nos principais métodos de seleção empregados no melhoramento de espécies perenes é importante considerar todas as covariâncias essenciais na adoção desses métodos. Assim, são relevantes as covariâncias:

$\operatorname{cov}(x_{ijk}, y)$, $\operatorname{cov}(\bar{x}_{i..}, y)$, $\operatorname{cov}[(x_{ijk} - \bar{x}_{ij.}), y]$ e $\operatorname{cov}(\bar{x}_{i..}, y_{pi})$, em que:

$\operatorname{cov}(x_{ijk}, y)$: covariância entre o indivíduo (x_{ijk}) como unidade de seleção e o indivíduo y da população melhorada (associada à seleção individual).

$\operatorname{cov}(\bar{x}_{i..}, y)$: covariância entre a média de família ($\bar{x}_{i..}$) como unidade de seleção e o indivíduo y da população melhorada (associada à seleção entre progênies).

$\operatorname{cov}[(x_{ijk} - \bar{x}_{ij.}), y]$: covariância entre o desvio do indivíduo (x_{ijk}) da média da família na parcela ($\bar{x}_{ij.}$) e o indivíduo y da população melhorada (associada à seleção dentro de progênies).

$\operatorname{cov}(\bar{x}_{i..}, y_{pi})$: covariância entre a média da família ($\bar{x}_{i..}$) como unidade de seleção e o indivíduo y_{pi} da população melhorada obtido através da

recombinação do genitor pi da progênie \bar{x}_i . (associada à seleção de genitores com base em teste de suas progênies).

Estas covariâncias estão associadas aos métodos de seleção: individual, entre progênies, dentro de progênies, entre e dentro de progênies, seleção pelo índice multiefeitos e seleção de genitores. Logicamente, esses métodos citados podem ser univariados (seleção baseada em apenas um caráter) ou multivariados (baseados em índices multicaracteres), de forma que os multiplicadores gerados a partir destas covariâncias poderão ser utilizados como ponderadores tanto de componentes de variância e/ou de covariância, dependendo se os métodos de seleção são univariados ou multivariados. A estimativa de componentes de variância e a predição de valores genéticos pelos métodos apresentados, sob sistema reprodutivo misto, podem ser realizados pelo software Selegen (Resende et al., 1994a).

As covariâncias (considerando o progresso genético imediato) para a seleção massal, entre progênies e seleção de genitores foram apresentadas por Wright & Cockerham (1985, 1986), enquanto para a seleção dentro de progênies, foram derivadas por Resende et al. (1995c). Na Tabela 18.1 são apresentados os multiplicadores dos componentes da variância genética, advindos da covariância unidade de seleção – população melhorada, para os métodos de seleção mencionados.

Tabela 18.1. Coeficientes dos componentes quadráticos, utilizados na covariância da expressão do progresso genético imediato, associado a diferentes métodos de seleção* com progênies de polinização aberta em populações de sistema reprodutivo misto.

Métodos	Coeficientes dos componentes de variância genética					
	σ_a^2	σ_d^2	D_1	D_2	H	H^2
de seleção						
Métodôs						
Individual	$\frac{2}{(2-S)}$	$\frac{S(1-S)}{(2-S)}$	$\frac{S(3+S)}{(2-S)}$	$\frac{S^2}{(2-S)}$	$\frac{S^2(1-S)}{(2-S)(4-S)}$	$\frac{2S^2(1-S)}{(2-S)^2(4-S)}$
Entre pro- gênies	$\frac{(1+S)^2}{2(2-S)}$	$\frac{S^2(1-S)}{2(2-S)}$	$\frac{S(1+S)^2}{(2-S)}$	$\frac{S^2(1+3S)}{4(2-S)}$	$\frac{S^3(1-S)}{2(2-S)(4-S)}$	$\frac{S^3(1-S)}{(2-S)^2(4-S)}$
Dentro de progênies	$\frac{4-(1-S)^2}{2(2-S)}$	$\frac{S(S-1)^2}{(2-S)}$	$\frac{S(-S^2-S+2)}{(2-S)}$	$\frac{S^2(3-3S)}{4(2-S)}$	$\frac{S^2(S^2-3S+2)}{2(2-S)(4-S)}$	$\frac{S^2(S^2-3S+2)}{(2-S)^2(4-S)}$
Seleção de genitores	$\frac{(1+S)}{(2-S)}$	$\frac{S^2(1-S)}{2(2-S)}$	$\frac{S(1+S)(3+S)}{2(2-S)}$	$\frac{S^2(1+3S)}{4(2-S)}$	$\frac{S^3(1-S)}{2(2-S)(4-S)}$	$\frac{S^3(1-S)}{(2-S)^2(4-S)}$
σ_g^2	$\frac{2}{(2-S)}$	$\frac{2(1-S)}{(2-S)}$	$\frac{4S}{(2-S)}$	$\frac{S}{(2-S)}$	$\frac{2S(1-S)}{(2-S)(4-S)}$	$\frac{4S(1-S)}{(2-S)^2(4-S)}$

* Considerou-se a seleção em ambos os sexos e número suficientemente grande de indivíduos por família.

Na obtenção dos coeficientes da Tabela 18.1, são relevantes as seguintes covariâncias:

(a) Covariância para a seleção massal ou individual

$$\text{cov}(x_{ijk}, y) = S \text{COV}(M, F_A) + 2(1-S) \text{COV}(M, F_C)$$

(b) Covariância para a seleção entre progênies

$$\text{cov}(\bar{x}_{i..}, y) = (1-S)^2 \text{COV}(I_C, I_C) + 2S(1-S) \text{COV}(I_A, I_C) + S^2 \text{COV}(I_A, I_A)$$

(c) Covariância para a seleção dentro de progênies

$$\text{cov}[(x_{ijk} - \bar{x}_{ij.}), y] = \text{cov}(x_{ijk}, y) - \text{cov}(\bar{x}_{i..}, y)$$

(d) Covariância para a seleção de genitores

$$\text{cov}(\bar{x}_{i..}, y_{pi}) = 2(1-S)^2 \text{COV}(I_C, I_C) + 2S(1-S) \text{COV}(I_A, I_C) + S(1-S) \text{COV}(I_C, I_A) + S^2 \text{COV}(I_A, I_A)$$

em que:

$\text{COV}(M, F_A)$: covariância entre mãe e filho obtido por autofecundação.

$\text{COV}(M, F_C)$: covariância entre mãe e filho obtido por cruzamento.

$\text{COV}(I_C, I_C)$: covariância entre irmãos originados por cruzamento.

$\text{COV}(I_A, I_C) = \text{COV}(I_C, I_A)$: covariância entre um indivíduo originado por cruzamento e seu irmão originado por autofecundação.

$\text{COV}(I_A, I_A)$: covariância entre irmãos originados por autofecundação.

Os conteúdos destas covariâncias entre parentes são dadas por:

$$\text{COV}(M, F_A) = (1+F) \sigma_a^2 + \frac{(1-F)}{2} \sigma_d^2 + \frac{(1+7F)}{2} D_1 + FD_2 + \frac{F(1-F)}{2(2+F)} H + \frac{F(1-F^2)}{2(2+F)} H^2;$$

$$\text{COV}(M, F_C) = \frac{(1-F)}{2} \sigma_a^2 + FD_1;$$

$$\text{COV}(I_C, I_C) = \frac{(1+F)}{4} \sigma_a^2 = \frac{1}{2(2-S)} \sigma_a^2;$$

$$\text{COV}(I_A, I_A) = (1+F) \sigma_a^2 + \frac{(1-F)}{4} \sigma_d^2 + (1+3F) D_1 + \frac{1+7F}{8} D_2 + \frac{F(1-F)}{4} H + \frac{F(1-F^2)}{4(2+F)} H^2;$$

$$= \frac{2}{2-S} \sigma_a^2 + \frac{(1-S)}{2(2-S)} \sigma_d^2 + \frac{2(1+S)}{2-S} D_1 + \frac{1+3S}{4(2-S)} D_2 + \frac{S(1-S)}{2(4-S)(2-S)} H + \frac{S(1-S)}{(4-S)(1-S)^2} H^2;$$

$$\text{COV}(I_A, I_C) = \frac{(1+F)}{2} \sigma_a^2 + \frac{(1+3F)}{4} D_1$$

A covariância entre irmãos germanos, sob sistema reprodutivo misto, equivale a

$$\text{COV}(IGM) = \frac{(1+F)}{2} \sigma_a^2 + \frac{(1+F)^2}{4} \sigma_d^2, \text{ em que } F = \frac{S}{2-S}.$$

18.4. Parâmetros Genéticos

Com vistas à obtenção de multiplicadores para a correção dos parâmetros genéticos obtidos a partir de progêniess maternais em espécies de sistema reprodutivo misto, Resende et al. (1995c) obtiveram os valores esperados dos parâmetros genéticos para três tipos de populações quanto ao grau de melhoramento, a saber: pouco melhorada, nível intermediário de melhoramento e melhorada, com freqüências alélicas médias dos alelos favoráveis em torno de 0,3; 0,5 e 0,7, respectivamente.

Para a obtenção dos componentes da variância genética, utilizou-se o mesmo procedimento adotado por Curnow & Baker (1968), Pereira (1989) e Fernandes (1990), em que as diferentes freqüências alélicas de uma determinada população se ajustam a uma distribuição beta. Assim, os valores das freqüências alélicas e as funções densidade de probabilidade das distribuições foram utilizadas em conjunto com as expressões (fornecidas por Cockerham & Weir, 1984) dos componentes da variação para se obterem os valores esperados dos parâmetros populacionais para um número qualquer n de locos.

A partir dos valores esperados de cada parâmetro, obtiveram-se os quocientes entre a variância genética aditiva e os demais componentes da variância genética para cada freqüência alélica e nível de dominância (grau médio de dominância) conforme Fernandes (1990). Também foram obtidos os quocientes entre a variância genética aditiva na freqüência alélica média de 0,5 e o mesmo parâmetro nas freqüências alélicas médias de 0,3 e 0,7 para diferentes graus médio de dominância.

A partir destes quocientes e da estimativa base da variância genética aditiva de 1,845, obtiveram-se os valores dos parâmetros utilizados. Supôs-se que a estimativa base de σ_a^2 é adequada a uma população com freqüência alélica intermediária ($p = 0,5$) e grau médio de dominância 1,0. Trabalhou-se com graus médio de dominância variando de 0,25 (dominância parcial) a 1,25 (sobredominância).

Merce destaque a obtenção do componente H^2 , o qual é dependente do número de locos controlando o caráter (que é desconhecido). Isto porque as expressões referentes aos demais componentes são função de n (que cancela nas relações dos parâmetros com a σ_a^2) e a expressão referente ao H^2 é função de n^2 . Entretanto, $H^2 = n \sigma_d^2$, de forma que estimando-se n e tendo-se σ_d^2 , obtém-se H^2 . Considerou-se $n = 100$ locos, e o parâmetro H^2 foi obtido para cada população por meio de $100 \sigma_d^2$. É importante ressaltar que a influência do valor de H^2 nos resultados foi, provavelmente, muito pequena, visto que os coeficientes associados ao mesmo são extremamente baixos em relação aos demais (Tabela 18.2.)

Na Tabela 18.2 são apresentados os parâmetros utilizados para cada uma das 15 populações.

Tabela 18.2. Valores dos componentes de variância genética utilizados para as 15 populações (combinações das freqüências alélicas médias – p e graus médio de dominância – g.m.d).

		g.m.d				
p		0,25	0,5	0,75	1,00	1,25
0,3	σ_a^2	2,056	2,343	2,611	2,844	3,043
	σ_d^2	0,045	0,174	0,377	0,632	0,925
	D ₁	-0,165	-0,360	-0,573	-0,798	-1,000
	D ₂	0,039	0,159	0,329	0,553	0,809
	H	0,045	0,174	0,377	0,632	0,925
	H ²	4,500	17,400	37,700	63,200	92,500
0,5	σ_a^2	1,845	1,845	1,845	1,845	1,845
	σ_d^2	0,049	0,191	0,412	0,692	1,000
	D ₁	-0,016	-0,064	-0,137	-0,230	-0,337
	D ₂	0,033	0,127	0,275	0,460	0,673
	H	0,049	0,191	0,412	0,692	1,000
	H ²	4,900	19,100	41,200	69,200	10,00
0,7	σ_a^2	1,474	1,210	0,981	0,789	0,645
	σ_d^2	0,045	0,174	0,377	0,632	0,925
	D ₁	0,126	0,207	0,243	0,237	0,197
	D ₂	0,039	0,153	0,329	0,553	0,809
	H	0,045	0,174	0,377	0,632	0,925
	H ²	4,500	17,400	37,700	63,200	92,500

Observa-se que o componente H apresenta valor idêntico ao de σ_d^2 , pois considerou-se o modelo com dois alelos por locos, conforme Wright & Cocke-rham (1985; 1986).

Os valores associados a H² foram muito superiores em relação aos de-mais componentes (Tabela 18.2), fato também verificado por Wright & Cocke-rham (1985; 1986), e que é teoricamente esperado, com grande número de lo-cos controlando o caráter. Os componentes H, σ_d^2 , H² e D₂ são simétricos em relação à freqüência alélica de 0,5, a qual representa os valores máximos para H, σ_d^2 e H² e mínimos para D₂. Dos novos componentes que surgem na pre-sença de endogamia, D₁ merece mais atenção, pois apresentou valores negati-vos nas freqüências alélicas de 0,3 e 0,5.

18.5. Efeitos nas Estimativas de Parâmetros Genéticos e nos Métodos de Seleção

A consideração do modelo completo, sob endogamia, nos programas de melhoramento aplicado, exige a estimação de todos os componentes da varia-ção genética, por meio da utilização de complexos delineamentos de cruzamen-

to. Estes delineamentos propostos, via de regra, exigem a obtenção e a avaliação de vários tipos de progênies (autofecundadas, meios-irmãos, polinização aberta, irmãos germanos) e avaliação dos genitores (Cockerham & Weir, 1984; Wright & Cockerham, 1985; 1986). Outras modificações foram propostas (Pereira, 1989; Crisóstomo, 1989), apresentando delineamentos mais simples, mas mesmo assim exigindo a avaliação de várias entidades genéticas nos experimentos de melhoramento. Estes últimos delineamentos propostos, todavia, não forneceram eficiência satisfatória para a estimativa dos componentes da variância genética (Crisóstomo, 1989; Pereira, 1989). Assim sendo, verifica-se que a adoção do modelo completo, com a estimativa de todos os parâmetros genéticos é, no momento, praticamente inviável na rotina de um programa de melhoramento de plantas perenes.

Para verificar a tendenciosidade na estimativa dos parâmetros genéticos, quando se consideram progênies de polinização aberta como de meios-irmãos, deve-se comparar as covariâncias entre a unidade de seleção e a população melhorada em ambos os casos, ou seja, quando se trabalha com progênies de meios-irmãos de uma população alógama e quando se trabalha com progênies de polinização aberta de uma população com sistema reprodutivo misto. Nas variâncias fenotípicas das unidades de seleção, nunca se incorrerá em erros, pois em ambos os casos são função dos componentes de variância entre progênies, entre parcelas e dentro de parcela, tal como sempre é usado.

Considerando, inicialmente, a seleção entre progênies, constata-se que em ambos os casos (progênies de polinização aberta e de meios irmãos) a covariância ($\text{cov}_{us, pm}$) entre a unidade de seleção e a população melhorada é exatamente igual à variância genética liberada entre progênies (σ_p^2), ou seja, para meios-irmãos legítimos $\text{COV}_{us, pm} = \sigma_p^2 = 1/4 \sigma_a^2$ e para progênies de polinização aberta:

$$\text{COV}_{us, pm} = \sigma_p^2 = \frac{(1+S)^2 \sigma_a^2}{2(2-S)} + \frac{S^2(1-S) \sigma_d^2}{(2-S)} + \frac{S(1+S)^2 D_1}{(2-S)} + \frac{S^2(1+3S) D_2}{4(2-S)} + \frac{S^3(1-S) H}{2(4-S)(2-S)} + \frac{S^3(1-S) H^2}{(4-S)(2-s)^2}$$

Esta última igualdade encontra-se em Wright & Cockerham (1986).

Assim sendo, não existe erro na seleção entre progênies e o componente de variância entre progênies pode ser usado diretamente na expressão de ganho, tal como se faz com meios-irmãos legítimos.

Para a seleção massal, a seleção de genitores e a seleção dentro de progênies, trabalhando-se com progênies de meios-irmãos, multiplica-se o componente de variância entre progênies por 4, 2 e 3, respectivamente, a fim de obter as respectivas covariâncias para os ganhos. No caso de progênies de polinização aberta, de populações com sistema reprodutivo misto, as relações entre as covariâncias unidade de seleção – população melhorada e a variância genética total liberada entre progênies são apresentadas na Tabela 18.3 para as 15 populações e taxas de autofecundação de 5% a 40%.

Tabela 18.3. Relações entre as covariâncias unidade de seleção – população melhorada e a variância genética entre progênies, para a seleção massal (M) e para a seleção de genitores com base em testes de progênies (TP). Quinze populações em função da freqüência alélica média (\bar{p}) e grau médio de dominância (gd).

Taxa de autofecundação: 5%.

\bar{p}						
0,3		0,5		0,7		
Gd	M	TP	M	TP	M	TP
0,25	3,64	1,90	3,63	1,90	3,62	1,90
0,50	3,65	1,91	3,65	1,90	3,63	1,89
0,75	3,67	1,91	3,67	1,90	3,67	1,89
1,00	3,69	1,91	3,70	1,90	3,73	1,89
1,25	3,71	1,91	3,74	1,91	3,84	1,89
Média	3,67	1,91	3,68	1,90	3,70	1,89

Média Geral: M = 3,68; TP = 1,90

Taxa de autofecundação: 10%.

Taxa de autofecundação: 15%.

\bar{p}						
0,3		0,5		0,7		
Gd	M	TP	M	TP	M	TP
0,25	3,06	1,75	3,05	1,74	3,03	1,73
0,50	3,11	1,75	3,11	1,74	3,09	1,71
0,75	3,18	1,75	3,21	1,73	3,24	1,69
1,00	3,26	1,75	3,33	1,72	3,49	1,65
1,25	3,34	1,75	3,47	1,72	3,85	1,61
Média	3,19	1,75	3,23	1,73	3,34	1,68
Média Geral: M = 3,25; TP = 1,72						

Taxa de autofecundação: 20%.

	\bar{p}					
	0,3		0,5		0,7	
Gd	M	TP	M	TP	M	TP
0,25	2,82	1,68	2,81	1,66	2,79	1,65
0,50	2,89	1,68	2,89	1,66	2,88	1,63
0,75	2,99	1,68	3,02	1,65	3,08	1,59
1,00	3,09	1,67	3,19	1,63	3,38	1,54
1,25	3,20	1,67	3,37	1,61	3,79	1,48
Média	3,00	1,68	3,06	1,64	3,18	1,58
Média Geral: M = 3,08; TP = 1,63						

Média Geral: M = 3,08; TP = 1,63

Taxa de autofecundação: 25%.

Gd	\bar{p}					
	0,3		0,5		0,7	
	M	TP	M	TP	M	TP
0,25	2,62	1,61	2,60	1,60	2,59	1,58
0,50	2,70	1,61	2,71	1,58	2,69	1,55
0,75	2,83	1,60	2,87	1,56	2,93	1,50
1,00	2,95	1,59	3,07	1,54	3,27	1,44
1,25	3,09	1,58	3,28	1,52	3,71	1,37
Média	2,84	1,60	2,91	1,56	3,04	1,49
Média geral: M = 2,93; TP = 1,55						

Média geral: M = 2,93; TP = 1,55

Taxa de autofecundação: 30%.

Média Geral: M = 2,76; TP = 1,47

Taxa de autofecundação: 35%.

Taxa de autofecundação: 40%.

Gd	\bar{p}					
	0,3		0,5		0,7	
	M	TP	M	TP	M	TP
0,25	2,11	1,43	2,10	1,42	2,08	1,40
0,50	2,24	1,42	2,25	1,39	2,23	1,36
0,75	2,41	1,40	2,46	1,36	2,52	1,29
1,00	2,58	1,38	2,69	1,32	2,84	1,23
1,25	2,74	1,36	2,92	1,28	3,19	1,17
Média	2,42	1,40	2,48	1,35	2,57	1,29
Média geral: M = 2,49; TP = 1,35						

Na Tabela 18.3 são apresentados os multiplicadores adequados da variância genética entre progênies, para as covariâncias associadas à seleção massal e à seleção de genitores. A covariância para a seleção dentro, com número grande de indivíduos por família equivale à covariância para a seleção massal menos a covariância para a seleção entre progênies. Dessa forma os multiplicadores (covariância dentro/covariância entre) adequados para a seleção dentro de progênies podem ser facilmente obtidos, subtraindo-se 1 dos valores para a seleção massal apresentados na Tabela 18.3.

Constata-se que, de maneira geral, considerar progênies de polinização aberta como progênies de meios-irmãos, conduz a superestimativas dos ganhos com a seleção massal, de genitores e dentro de progênies, conforme pode ser observado comparando-se os valores que deveriam ser usados (Tabela 18.3), com aqueles tradicionalmente usados: 4, 2 e 3, respectivamente. E esta superestimação é tanto maior quanto maior for a taxa de autofecundação. Embora não utilizando o modelo genético adequado, Squillace (1974) também concluiu que, de maneira geral, ocorrem superestimativas dos ganhos para a seleção massal.

Estas superestimativas são verdadeiras também para a seleção pelo índice multiefeitos (uma vez que ele depende da herdabilidade dentro de progênies), e para os métodos multivariados (índice de seleção) associados à seleção massal, dentro de progênies, índice multiefeitos e seleção de genitores.

Observa-se que os multiplicadores encontrados para cada taxa de autofecundação variam em função do grau médio de dominância e, em menor grau, do nível de melhoramento das populações. Assim sendo, a utilização desses multiplicadores, de forma eficiente, nos programas de melhoramento aplicado dependem, sobretudo, do conhecimento do grau médio de dominância do caráter.

De qualquer forma, constata-se, pelos valores da média geral dos multiplicadores para cada taxa de autofecundação, o comportamento claro de diminuição desses multiplicadores, em função do aumento da taxa de autofecundação (Tabela 18.3). Assim sendo, conhecendo-se a taxa de autofecundação, uma opção razoável de utilização desses multiplicadores seria optar pelos valores mais baixos no geral ou para um grau médio de dominância aproximado, e trabalhar, assim, com subestimativas e não superestimativas de ganhos genéticos.

No melhoramento florestal, vários autores têm proposto algumas formas de correção das estimativas dos parâmetros genéticos sob endogamia. Namkoong (1966) relata que a variância genética aditiva pode ser estimada sob endogamia a partir de progênies de polinização aberta somente para casos de ausência de dominância. De fato, neste caso, a variância genética entre progênies somente é função da variância genética aditiva, não participando os componentes σ_d^2 , D_1 , D_2 , H e H^2 . Assim, a correção proposta por esse autor, com base apenas no coeficiente de endogamia, somente é válida para o caso especificado. Entretanto, na maioria das vezes, não se têm bases reais para admitir esta suposição.

A correção proposta por Squillace (1974) baseia-se na determinação de correlações genéticas médias entre os indivíduos de uma família de polinização aberta, partindo do fato de que uma família de polinização aberta, numa população com número finito de machos polinizadores, apresenta indivíduos cruzados e autofecundados. Assim, essa família é composta por diversos tipos de parentes tais como irmãos S_1 , irmãos germanos, meios-irmãos e irmãos cruzados e autofecundados. Dessa forma, o autor determinou, para vários graus de autofecundação, as correlações genéticas médias (ρ_a) entre os indivíduos de famílias de polinização aberta. O autor propôs a expressão $\sigma_a^2 = \sigma_p^2 / \rho_a$, para correção da covariância para a herdabilidade associada à seleção massal. Entretanto, o uso dessa expressão pressupõe que, com autofecundação, a variância genética entre progênies é função apenas da variância genética aditiva, suposição essa falsa quando existe dominância conforme corroborado pela expressão da variância entre progênies já apresentada anteriormente.

Na Tabela 18.4, são apresentados os multiplicadores da variância genética entre progênies, com vistas à determinação da covariância para a herdabilidade associada à seleção massal, com base em duas aproximações, a saber: (a) valores extraídos a partir das covariâncias apresentadas por Squillace (1974); b) valores obtidos a partir da relação coeficiente da variância aditiva associada à herdabilidade da seleção massal e coeficiente da variância aditiva liberada entre progênies (usando o modelo completo).

Tabela 18.4. Multiplicadores da variância genética entre progênies, para a determinação da covariância adequada à herdabilidade individual, obtidos com base em duas aproximações: a) Squillace (1974); b) relação entre o coeficiente da variância aditiva associada à herdabilidade da seleção individual e o coeficiente da variância aditiva liberada entre progênies [$4/(1+S)^2$].

Taxa de autofecundação	Coeficientes	
	a	b
5%	3,6	3,6
10%	3,4	3,3
15%	3,2	3,0
20%	3,0	2,8
25%	2,9	2,6
30%	2,7	2,4
35%	2,6	2,2
40%	2,4	2,0

As duas metodologias apresentadas baseiam-se apenas nos coeficientes da variância genética associado aos efeitos aditivos. O método b é mais adequado pois considera as covariâncias apresentadas por Cockerham & Weir (1984) e também a fração da variância aditiva explorada pela seleção massal, que com endogamia é sempre maior que a variância genética aditiva da população não endógama. Assim, considerando apenas variância genética aditiva, verifica-se que a metodologia de Squillace (1974) superestima os ganhos genéticos com seleção, tanto mais quanto maior for a taxa de autofecundação (Tabela 18.4).

Na Tabela 18.5 são apresentados, para o grau de dominância 1 e considerando $\bar{p} = 0,5$, os coeficientes (extraídos da Tabela 18.3.) a serem usados na multiplicação da variância genética entre progênies visando obter as covariâncias para a seleção massal, de genitores e dentro de progênies. São apresentados, também, esses mesmos coeficientes, porém considerando somente as relações entre a variância genética aditiva (modelo aditivo, sem dominância).

Tabela 18.5. Coeficientes a serem utilizados na multiplicação da variância genética entre progênies, visando obter as covariâncias para a seleção massal (M), de genitores (G) e dentro de progênies (D), obtidos a partir dos modelos de dominância completa ($\bar{p} = 0,5$) e completamente aditivo.

Coeficiente de autofecundação (S)	Modelo completo			Somente variância aditiva		
	M	G	D	M*	G**	D***
autofecundação (S)	3,7	1,9	2,7	3,6	1,9	2,6
10%	3,6	1,8	2,6	3,3	1,8	2,3
15%	3,3	1,7	2,3	3,0	1,7	2,0
20%	3,2	1,6	2,2	2,8	1,7	1,8
25%	3,1	1,5	2,1	2,6	1,6	1,6
30%	2,9	1,5	1,9	2,4	1,5	1,4
35%	2,8	1,4	1,8	2,2	1,5	1,2
40%	2,7	1,3	1,7	2,0	1,4	1,0

$$* M = 4 / (1 + S)^2$$

$$** G = 2 (1 + S) / (1 + S)^2$$

$$*** D = 4 / (1 + S)^2 - 1$$

Verifica-se que os multiplicadores associados ao modelo com dominância completa são sempre iguais ou superiores àqueles associados ao modelo completamente aditivo. Assim, adotando os multiplicadores do modelo completamente aditivo, incorre-se, na pior das hipóteses, em subestimativas do ganho genético. Assim, adotar generalizadamente os multiplicadores $4 / (1 + S)^2$, $2 (1 + S) / (1 + S)^2$; 1 ; $4 (1 + S)^2 - 1$ para as covariâncias associadas à seleção massal, seleção de genitores, seleção entre progênies e seleção dentro de progênies, respectivamente, é uma estratégia conservadora e mais segura.

Conhecendo-se aproximadamente o grau médio de dominância do caráter, pode-se usar com vantagens os coeficientes de correlação genética entre parentes ou os multiplicadores (entre parênteses) apresentados na Tabela 18.6, para cada grau de dominância e taxa de autofecundação.

Tabela 18.6. Correlações genéticas entre parentes (ρ_g) e multiplicadores (m_p) da variância genética entre progênies maternais, para vários graus de dominância (gd) e taxas de autofecundação (S).

S(%)	gd					
	0,00 $\rho_g(m_p)$	0,25 $\rho_g(m_p)$	0,50 $\rho_g(m_p)$	0,75 $\rho_g(m_p)$	1,00 $\rho_g(m_p)$	1,25 $\rho_g(m_p)$
0%	0,2500 (4,00)	0,25 (4,0)	0,25 (4,0)	0,25 (4,0)	0,25 (4,0)	0,25 (4,0)
5%	0,2756 (3,63)	0,28 (3,6)	0,27 (3,7)	0,27 (3,7)	0,27 (3,7)	0,26 (3,8)
10%	0,3025 (3,31)	0,30 (3,3)	0,30 (3,3)	0,29 (3,4)	0,28 (3,6)	0,27 (3,7)
15%	0,3306 (3,02)	0,33 (3,0)	0,32 (3,1)	0,31 (3,2)	0,30 (3,3)	0,28 (3,6)
20%	0,3600 (2,78)	0,36 (2,8)	0,35 (2,9)	0,33 (3,0)	0,31 (3,2)	0,29 (3,4)
25%	0,3906 (2,56)	0,38 (2,6)	0,37 (2,7)	0,35 (2,9)	0,32 (3,1)	0,30 (3,3)
30%	0,4225 (2,37)	0,42 (2,4)	0,40 (2,5)	0,37 (2,7)	0,34 (2,9)	0,31 (3,2)
35%	0,4556 (2,19)	0,45 (2,2)	0,42 (2,4)	0,39 (2,6)	0,36 (2,8)	0,32 (3,1)
40%	0,4900 (2,04)	0,48 (2,1)	0,45 (2,2)	0,41 (2,4)	0,37 (2,7)	0,34 (2,9)

Pela Tabela 18.6, constata-se que, quando a taxa de autofecundação for maior que 5%, faz-se necessária a correção das estimativas dos parâmetros genéticos.

De maneira geral, fica claro que a variância genética aditiva é preponderante na determinação das relações apresentadas e discutidas, sobretudo em função dos maiores coeficientes associados a este componente (Tabela 18.1). A utilização dos coeficientes obtidos com base apenas na variância aditiva parece adequada, pois eles dependem apenas da taxa de autofecundação, independentemente do grau médio de dominância, do nível de melhoramento das populações e do número de locos que controla o caráter. Na pior das hipóteses (se o grau médio de dominância for extremamente alto), trabalha-se com subestimativas e não superestimativas de ganhos genéticos. Outro aspecto interessante é que estes coeficientes poderão ser utilizados generalizadamente para quaisquer caracteres, em quaisquer espécies. Poderão ser utilizados, também, para cálculos de ganhos genéticos permanentes. Nestes casos, os coeficientes da variâ-

cia genética aditiva permanecem os mesmos verificados para os progressos imediatos, os componentes D_1 e D_2 apresentam menores coeficientes e os componentes H , H^2 e σ_d^2 não são explorados pela seleção.

Fica patente a necessidade de adotar alguma técnica de correção nas estimativas dos parâmetros genéticos em populações com sistema reprodutivo misto. No melhoramento florestal, freqüentemente (Griffin & Coterill, 1980; Völker et al., 1990; Borralho et al., 1992) têm sido adotadas correções com base na metodologia de Squillace (1974).

Os resultados apresentados, baseados apenas na variância genética aditiva, são mais acurados e subestimados em relação aos apresentados por Squillace (1974) e, portanto, poderão ser usados vantajosamente. Os multiplicadores relatados, para os graus médio de dominância 1 e 0, foram utilizados no estudo de parâmetros genéticos e métodos de seleção em seringueira por Costa et al. (2000).

Os resultados apresentados representam subsídios importantes para a seleção em populações com sistema reprodutivo misto. Entretanto, torna-se necessária a determinação precisa das taxas de autofecundação para as diferentes populações nos diferentes locais. Isto pode ser realizado eficientemente por meio da técnica da eletroforese de isoenzimas, conforme descrito por Ritland & Jain (1981) e Muona & Harju (1989).

18.6. Comparação entre Tipos de Progênies para Maximização do Ganhos Genéticos

Neste contexto, trabalho relevante foi realizado por Pereira (1989), comparando progênies de meios-irmãos, maternais (polinização aberta) e S_1 's. De maneira geral, tal autor concluiu pela superioridade da seleção com S_1 , seguida pela seleção com progênies maternais e meios-irmãos, nesta ordem. Esse autor concluiu, também que a vantagem de S_1 sobre maternais diminuiu com o aumento da taxa de autofecundação. Estes resultados são coerentes com os apresentados no capítulo sobre comparação de S_1 e meios-irmãos em espécies alógamas, e decorrem, sobretudo, da maior variação genética expressa entre progênies endógamas.

Outra modalidade de seleção poderia basear-se em um índice que combine as informações das progênies maternais e autofecundadas. Este índice pode ser construído tendo por base dois objetivos alternativos: 1) o melhoramento populacional e a simultânea redução da depressão endogâmica; 2) apenas o melhoramento populacional.

No primeiro caso, em termos de componentes de médias, deve-se maximizar o objetivo $O = (g - D)$, em que g refere-se ao valor genotípico médio. Maximizar $(g - D)$ significa maximizar g e minimizar D . A depressão D pode ser estimada a partir da diferença entre médias de progênies com diferentes coeficientes de endogamia. Assim, D pode ser estimada por:

- (i) diferença entre médias de progênies cruzadas (C, com $F = 0$) e autofecundadas (S_1 , com $F = 1/2$) em populações alógamas, por $(1/2) D = C - S_1$.
- (ii) diferença entre médias de progênies cruzadas em populações alógamas (C, com $F = 0$) e médias de linhagens (L, com $F = 1$), por $D = C - L$.
- (iii) diferença entre médias de progênies maternas (M, com $F = S/(2-S)$) e autofecundadas (S^* , com $F = (1/2)(1+F_p) = 1/(2-S)$) em populações com sistema misto de reprodução, por $[1/(2-S) - S/(2-S)] D = M - S^*$.

Na situação (iii), $F_p = S/(2-S)$ e refere-se ao coeficiente de endogamia do genitor, em que S é a taxa de autofecundação. Em tal caso, D é estimado por $[(1-S)/(2-S)] D = (M - S^*)$, que equivale a $D = (2 - S) / (1 - S) (M - S^*)$.

Retornando à questão da maximização, deseja-se, então, maximizar $O = (g - D)$, ou seja, maximizar $O = [g - [(2 - S) / (1 - S) (M - S^*)]]$. Verifica-se que maximizar O significa aumentar g e reduzir $(M - S)$, ou seja, melhorar a população e reduzir a depressão endogâmica. Sendo a seleção baseada no índice $I = b_1 M + b_2 S^*$, os pesos b_1 e b_2 devem ser determinados de forma que a correlação entre I e O seja máxima. Esta maximização conduz ao seguinte sistema matricial:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sigma_{F_M}^2 & Cov_g(M, S^*) \\ Cov_g(M, S^*) & \sigma_{F_{S^*}}^2 \end{bmatrix}^{-1} \begin{bmatrix} Cov[g - (M - S^*) \frac{2-S}{1-S}, M] \\ Cov[g - (M - S^*) \frac{2-S}{1-S}, S^*] \end{bmatrix}, \text{ em que:}$$

$\sigma_{F_M}^2$: variância fenotípica ao nível de médias das progênies M, prontamente estimável dos dados experimentais.

$\sigma_{F_{S^*}}^2$: variância fenotípica ao nível de médias das progênies S^* , prontamente estimável dos dados experimentais.

$Cov_g(M, S^*)$: covariância genética entre progênies M e S^* , quantidade prontamente estimável via análise de covariância entre progênies maternas e autofecundadas.

As covariâncias entre o objetivo O da seleção e as médias das progênies M e S^* são dadas por:

$$Cov[g - (M - S^*) \frac{2-S}{1-S}, M] = \sigma_{g(M)}^2 - \frac{2-S}{1-S} \sigma_{F_M}^2 + \frac{2-S}{1-S} Cov_g(M, S^*)$$

$$Cov[g - (M - S^*) \frac{2-S}{1-S}, S^*] = \sigma_{g(S^*)}^2 - \frac{2-S}{1-S} Cov_g(M, S^*) + \frac{2-S}{1-S} \sigma_{F_{S^*}}^2,$$

considerando a avaliação das progênies M e S em diferentes ambientes (blocos). Caso contrário,

$\text{Cov}(M, S^*)$ deve ser ao nível fenotípico e não genético.

Os componentes $\sigma_{g(M)}^2$ e $\sigma_{g(S^*)}^2$ referem-se às variâncias genéticas entre progêneres M e S^* e são prontamente estimáveis via análise de variância.

O sistema matricial pode ser reescrito como

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} C^{11} & C^{12} \\ C^{12} & C^{22} \end{bmatrix} \begin{bmatrix} [\text{Cov}(g-D), M] \\ [\text{Cov}(g-D), S^*] \end{bmatrix}, \text{ em que:}$$

$$C^{11} = \frac{\sigma_{F_{S^*}}^2}{\sigma_{F_M}^2 \sigma_{F_{S^*}}^2 - \text{Cov}_g(M, S^*)^2}; \quad C^{12} = \frac{-\text{Cov}_g(M, S^*)}{\sigma_{F_M}^2 \sigma_{F_{S^*}}^2 - \text{Cov}_g(M, S^*)^2}; \quad C^{22} = \frac{\sigma_{F_M}^2}{\sigma_{F_M}^2 \sigma_{F_{S^*}}^2 - \text{Cov}_g(M, S^*)^2}.$$

Assim:

$$\begin{aligned} b_1 &= C^{11}[\text{Cov}(g-D), M] + C^{12}[\text{Cov}(g-D), S^*] \\ b_2 &= C^{12}[\text{Cov}(g-D), M] + C^{22}[\text{Cov}(g-D), S^*] \end{aligned}$$

Este índice pode ser comparado com um índice I^* , visando apenas ao melhoramento do valor genético médio da população independentemente da depressão endogâmica, dado por $I^* = b_3 M + b_4 S^*$, em que:

$$\begin{bmatrix} b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} C^{11} & C^{12} \\ C^{12} & C^{22} \end{bmatrix} \begin{bmatrix} [\text{Cov}(g, M)] \\ [\text{Cov}(g, S^*)] \end{bmatrix}$$

$$\begin{aligned} \text{Cov}(g, M) &= \sigma_{g(M)}^2 \\ \text{Cov}(g, S^*) &= \sigma_{g(S^*)}^2 \end{aligned}$$

A correlação entre os índices I e I^* provavelmente será alta, significando que o melhoramento populacional já contribui para reduzir a depressão endogâmica.

18.7. Tamanho Efetivo em Populações com Sistema Reproutivo Misto

Com base no conceito de tamanho efetivo de variância (N_e), são derivadas, a seguir, expressões adequadas para o cômputo de N_e , em situações de representação desbalanceada das progêneres em termos do número de indivíduos em cada uma delas, situação esta muito comum no melhoramento e na conservação genética de plantas perenes.

A expressão geral do tamanho efetivo (N_e) equivale a (Crow & Kimura, 1970):

$$N_e = \frac{N \bar{k}}{\frac{\sigma_k^2}{\bar{k}} (1+\alpha) + (1-\alpha)} \quad (1)$$

em que:

N : número de plantas na população.

\bar{k} : número médio de gametas contribuídos pelas N plantas.

σ_k^2 : variância do número k de gametas contribuídos pelas N plantas.

α : desvio da população de genitores em relação ao equilíbrio de Hardy-Weinberg.

Para populações grandes, α equivale a zero para o caso de alogamia e $\alpha = F = S/(2-S)$ para o caso de sistema reprodutivo misto, em que S é a taxa de autofecundação.

Sendo f o número de matrizes amostradas e \bar{n} o número médio de indivíduos tomados por matriz, $\bar{k} = (2f\bar{n})/Ne$ para populações com sistema reprodutivo misto ($\alpha = S/(2-S)$), substituindo \bar{k} e α em (1), é dado por:

$$N_e = \frac{2(2-S) f^2 \bar{n}^2}{N \sigma_k^2 + 2(1-S) f \bar{n}} \quad (2)$$

Supondo N constante através das gerações, a expressão (2) reduz a

$$N_e = \frac{2(2-S) N}{\sigma_k^2 + 2(1-S)} = \frac{2(2-S) f \bar{n}}{\sigma_k^2 + 2(1-S)} \quad (3)$$

Assim, para cômputo do N_e , é necessário conhecer f , \bar{n} , S e σ_k^2 . Os valores de f , \bar{n} , e S são normalmente conhecidos na prática, de forma que é necessário detalhar apenas a forma de obtenção de σ_k^2 , a variância do número de gametas contribuídos.

No caso particular em que os números de descendentes autofecundados e cruzados são independentemente distribuídos como uma Poisson, $\sigma_k^2 = 4S + 2(1-S) = 2 + 2S$ e a equação (3) reduz-se a

$$N_e = \frac{N}{1+\alpha} = \frac{N (2-S)}{2} = \frac{f \bar{n} (2-S)}{2} \quad (4)$$

Por esta última expressão percebe-se que, para um tamanho populacional constante através das gerações, a ocorrência de autofecundação reduz o tamanho efetivo.

Entretanto, assumir distribuição de Poisson enfraquece o resultado. Um tratamento mais genérico e realístico para obtenção de σ_k^2 é apresentado a seguir.

Considere o seguinte esquema ilustrativo:

Matrizes	n_i	Gametas femininos		Gametas masculinos	
		\otimes	#	\otimes	#
1	n_1	Sn_1	$(1-S) n_1$	Sn_1	k_1
2	n_2	Sn_2	$(1-S) n_2$	Sn_2	k_2
3	n_3	Sn_3	$(1-S) n_3$	Sn_3	k_3
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
F	n_f	Sn_f	$(1-S)n_f$	Sn_f	k_f
⋮	⋮	⋮	⋮	⋮	k_{f+1}
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	k_M
⋮	⋮	⋮	⋮	⋮	0
N	0	0	0	0	0

em que:

f : número de matrizes (geração prévia) selecionadas.

n_i : número de indivíduos selecionados por matriz (progênie).

\otimes : autofecundação.

#: cruzamento.

k_i : número de gametas masculinos contribuídos via cruzamento. Não existe qualquer tipo de seleção (dependente da seleção das f matrizes via seus descendentes) sobre estes gametas.

S : taxa de autofecundação.

Denotando-se:

g : gametas masculinos.

d : gametas femininos.

z : gametas contribuídos por autofecundação.

x : gametas femininos contribuídos via cruzamento.

y : gametas masculinos contribuídos via cruzamento.

Tem-se que:

$$k = \sum \text{gametas} = d + g$$

$$d = x + z$$

$$g = z + y$$

Assim,

$$\sigma_k^2 = \sigma_d^2 + \sigma_g^2 + 2Cov_{dg}$$

$$\sigma_d^2 = \sigma_x^2 + \sigma_z^2 + 2Cov_{xz}$$

$$\sigma_g^2 = \sigma_z^2 + \sigma_y^2 + 2Cov_{zy}$$

$2Cov_{dg} = 2[Cov_{xy} + Cov_{xz} + Cov_{yz}] = 2[Cov_{xy} + Cov_{xz} + V_z + Cov_{zy}]$ e portanto:

$$\sigma_k^2 = \sigma_x^2 + \sigma_z^2 + 2Cov_{xz} + \sigma_z^2 + \sigma_y^2 + 2Cov_{zy} + 2Cov_{xy} + 2Cov_{xz} + 2\sigma_z^2 + 2Cov_{zy}$$

$$\sigma_k^2 = \sigma_x^2 + 4\sigma_z^2 + \sigma_y^2 + 4Cov_{xz} + 4Cov_{zy} + 2Cov_{xy} \quad (5)$$

No grupo selecionado, $Cov_{xy} = Cov_{zy} = 0$, pois y_i do genitor i independe de seu x_i e z_i .

Assim:

$$\sigma_k^2 = \sigma_x^2 + 4\sigma_z^2 + \sigma_y^2 + 4Cov_{xz} \quad (6)$$

É necessário então, estimadores para σ_x^2 , σ_z^2 , σ_y^2 e Cov_{xz} . Denotando-se esperança por E e variância por V, tem-se:

Distribuição de y_i : Binomial

Dentro do grupo dos M genitores masculinos selecionados, a distribuição de y_i é binomial. Assim,

$$E(y_i) = \frac{\sum_i (1-S) n_i}{M} \text{ e}$$

$V(y_i) = \sum_i (1-S) n_i \frac{1}{M} (1 - \frac{1}{M})$, que é a variância de uma variável binomial. Com M relativamente alto (>25), $(1-1/M) \approx 1$ e a distribuição de y_i aproxima uma Poisson com média e variância dados por

$$E(y_i) = V(y_i) = \frac{\sum_i (1-S) n_i}{M} = (1-S) \frac{\sum_i n_i}{M} = \frac{(1-S)f \bar{n}}{M} = \sigma_y^2 \quad (7)$$

Distribuição de z_i : Binomial Condisional a n_i

$$E(z_i / n_i) = S n_i$$

$$V(z_i / n_i) = S(1-S) n_i$$

$$\begin{aligned} V(z_i) &= E[V(z_i / n_i)] + V[E(z_i / n_i)] \\ &= E[S(1-S) n_i] + V[S n_i] \end{aligned} \quad (8)$$

$$= S(1-S) E(n_i) + S^2 V(n_i)$$

$$= S(1-S) \bar{n} + S^2 \sigma_n^2 = \sigma_z^2$$

Distribuição de x_i : Binomial Condicional a n_i

$$\begin{aligned} E(x_i / n_i) &= (1-S) n_i \\ V(x_i / n_i) &= S(1-S) n_i \\ V(x_i) &= E[V(x_i / n_i)] + V[E(x_i / n_i)] \\ &= S(1-S) \bar{n} + (1-S)^2 \sigma_n^2 = \sigma_x^2 \end{aligned} \quad (9)$$

Obtenção da Cov (z_i, x_i)

$$n = X + Z$$

$$\sigma_n^2 = \sigma_x^2 + \sigma_z^2 + 2Cov_{xz}. \text{ Usando-se (8) e (9) tem-se:}$$

$$\sigma_n^2 = S(1-S) \bar{n} + S^2 \sigma_n^2 + S(1-S) \bar{n} + (1-S)^2 \sigma_n^2 + 2Cov_{xz}, \text{ e, portanto,}$$

$$Cov_{xz} = S(1-S)[\sigma_n^2 - \bar{n}] \quad (10)$$

É necessário agora computar os componentes de σ_k^2 sobre todo o grupo de matrizes, selecionadas e não selecionadas.

Assim, por exemplo, σ_x^2 deve ser decomposto em $\sigma_x^2 = \sigma_{xse}^2 + \sigma_{xnse}^2 + \sigma_{x(se-nse)}^2$, em que **se** denota o grupo selecionado e **nse** denota o grupo não selecionado.

O diagrama seguinte ilustra a questão:

Lado feminino			
Matrizes	\otimes	#	
1	Z_1	X_1	$n_1 = Z_1 + X_1$
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
F	Z_f	X_f	$n_f = Z_f + X_f$
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
N	0	0	0

De acordo com o diagrama, tem-se:

$$\sigma_{xse}^2 = \frac{f}{N} \sigma_x^2; \sigma_{xnse}^2 = 0; \sigma_{x(se-nse)}^2 = \frac{f}{N} \left(1 - \frac{f}{N}\right) \bar{x}_{se}^2,$$

em que $\bar{x}_{se}^2 = (1-S) \bar{n}$.

Assim, σ_x^2 sobre todo o grupo das N matrizes pode ser computado em função de σ_x^2 e \bar{n} no grupo selecionado. Considerações análogas podem ser feitas para σ_z^2 e Cov_{xz} . Dessa forma, para o grupo das N matrizes originais, tem-se que:

$$\sigma_{x_N}^2 = \frac{f}{N} [S(1-S) \bar{n} + (1-S)^2 \sigma_n^2] + \frac{f}{N} (1 - \frac{f}{N}) [(1-S)^2 \bar{n}^2];$$

$$\sigma_{z_N}^2 = \frac{f}{N} [S(1-S) \bar{n} + S^2 \sigma_n^2] + \frac{f}{N} (1 - \frac{f}{N}) [S^2 \bar{n}^2];$$

$$Cov_{xz_N} = \frac{f}{N} [S(1-S) (\sigma_n^2 - \bar{n})] + \frac{f}{N} (1 - \frac{f}{N}) [S(1-S) \bar{n}^2];$$

$$\sigma_{y_N}^2 = \frac{M}{N} \left[\frac{f}{M} [(1-S) \bar{n}] \right] + \frac{M}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [(1-S) \bar{n}] \right]^2$$

Estas expressões são obtidas substituindo-se os estimadores de σ_x^2 , σ_z^2 , Cov_{xz} e σ_y^2 apresentados nas expressões (7), (8), (9) e (10).

Uma vez que $Cov_{xy} = Cov_{zy} = 0$ no grupo das f matrizes selecionadas, tem-se que, no conjunto das N plantas, Cov_{zy_N} e Cov_{xy_N} são dados por:

$$Cov_{zy_N} = \frac{f}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [S(1-S) \bar{n}^2] \right]$$

$$Cov_{xy_N} = \frac{f}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [(1-S)^2 \bar{n}^2] \right].$$

Torna-se claro que σ_n^2 e \bar{n} referem-se à variância e à média do número de indivíduos selecionados por matriz ou progénie, ou seja, ao grupo selecionado.

Voltando à expressão (5), agora no grupo total das N plantas, tem-se que:

$$\begin{aligned} \sigma_k^2 &= \sigma_{x_N}^2 + 4\sigma_{z_N}^2 + \sigma_{y_N}^2 + 4Cov_{xz_N} + 4Cov_{zy_N} + 2Cov_{xy_N} \\ &= \sigma_{d_N}^2 + \sigma_{g_N}^2 + 2Cov_{dg_N}; \end{aligned}$$

$$\sigma_{d_N}^2 = \sigma_{x_N}^2 + \sigma_{z_N}^2 + 2Cov_{xz_N} = \frac{f}{N} \sigma_n^2 + \frac{f}{N} (1 - \frac{f}{N}) \bar{n}^2;$$

$$\begin{aligned} \sigma_{g_N}^2 &= \sigma_{y_N}^2 + \sigma_{z_N}^2 + 2Cov_{zy_N} = \\ &= \frac{f}{N} [S^2 \sigma_n^2 + S(1-S) \bar{n}] + \frac{f}{N} (1 - \frac{f}{N}) (S^2 \bar{n}^2) + \frac{M}{N} [\frac{f}{M} (1-S) \bar{n}] + \frac{M}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [(1-S) \bar{n}] \right]^2 + 2 \frac{f}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [S(1-S) \bar{n}^2] \right] \end{aligned}$$

$$\begin{aligned} 2Cov_{dg_N} &= 2(\sigma_{z_N}^2 + Cov_{xz_N} + Cov_{xy_N} + Cov_{zy_N}) \\ &= \frac{f}{N} (2S\sigma_n^2) + \frac{f}{N} (1 - \frac{f}{N}) (2S\bar{n}^2) + 2 \frac{f}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [(1-S)^2 \bar{n}^2] \right] + 2 \frac{f}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [S(1-S) \bar{n}^2] \right] \end{aligned}$$

Assim:

$$\begin{aligned} \sigma_k^2 &= \sigma_{d_N}^2 + \sigma_{g_N}^2 + 2Cov_{dg_N} = \\ &= \frac{f}{N} [(1+S)^2 \sigma_n^2 + S(1-S) \bar{n}] + \frac{f}{N} (1 - \frac{f}{N}) [(1+S)^2 \bar{n}^2] + \frac{M}{N} [\frac{f}{M} (1-S) \bar{n}] + \frac{M}{N} (1 - \frac{M}{N}) \left[\frac{f}{M} [(1-S) \bar{n}] \right]^2 + \frac{f}{N} (1 - \frac{M}{N}) \left[2 \frac{f}{M} \bar{n}^2 [S(1-S) + (1-S)^2 + S(1-S)] \right] \end{aligned} \tag{11}$$

Considerando $(1 - \frac{f}{N}) \approx 1$ e $(1 - \frac{M}{N}) \approx 1$, tem-se:

$$\sigma_k^2 = \frac{f}{N}[(1+S)^2 \sigma_n^2 + S(1-S) \bar{n} + (1+S)^2 \bar{n}^2 + 2\frac{f}{M} \bar{n}^2[2S(1-S) + (1-S)^2]] + \frac{M}{N} \left[\frac{f}{M}(1-S) \bar{n} + [\frac{f}{M}(1-S) \bar{n}]^2 \right]$$

Substituindo σ_k^2 na expressão (2), obtém-se:

$$N_e = \frac{2(2-S) f \bar{n}}{(1+S)^2 \frac{\sigma_n^2}{\bar{n}} + [(1+S)^2 + \frac{f}{M}(3-2S-S^2)]\bar{n} + (3-2S-S^2)}$$

Considerando um grande número M de genitores masculinos, a expressão (12) é simplificada para

$$N_e = \frac{2(2-S) f \bar{n}}{(1+S)^2 \frac{\sigma_n^2}{\bar{n}} + (1+S)^2 \bar{n} + (3-2S-S^2)} \quad (13)$$

A expressão (13) é genérica e pode ser utilizada para plantas alógamas e com sistema reprodutivo misto. Para plantas alógamas $S = 0$ e, portanto, substituindo-se $S = 0$ na expressão (13) obtém-se:

$$N_e = \frac{4 f \bar{n}}{\frac{\sigma_n^2}{\bar{n}} + \bar{n} + 3}, \text{ que é a expressão do } N_e \text{ para plantas alógamas, empregada}$$

em vários outros capítulos deste livro. (14)

Em resumo, a expressão (13) é adequada para o cômputo do tamanho efetivo em populações com sistema reprodutivo, sob seleção individual, pelo índice multiefeitos ou pelo procedimento BLUP, a partir da avaliação de progênie de polinização aberta. Para o cômputo do N_e é necessário conhecer a taxa de autofecundação, o número de famílias selecionadas, o número médio de indivíduos selecionados e a variância do número de indivíduos selecionados por progénie. Tal N_e pode ser estimado pelo software Selegen (Resende et al., 1994a).

A expressão (12) pode ser aplicada a espécies autógamas, bastando fazer $M = f$ e substituir S por 1. Assim, obtém-se:

$$N_e = \frac{2(2-S) f \bar{n}}{(1+S)^2 \frac{\sigma_n^2}{\bar{n}} + 4\bar{n} + (3-2S-S^2)} = \frac{2f \bar{n}}{4 \frac{\sigma_n^2}{\bar{n}} + 4\bar{n}} = \frac{0,5f \bar{n}}{\frac{\sigma_n^2}{\bar{n}} + \bar{n}} \quad (15)$$

Considerando amostragem de sementes no grupo das f plantas femininas, ou seja, distribuição de Poisson, tem-se $\sigma_n^2 = \bar{n}$ e a expressão (15) passa a ser dada por

$$N_e = \frac{2f \bar{n}}{4 + 4\bar{n}} = \frac{0,5f \bar{n}}{1 + \bar{n}} \quad (16)$$

Com controle gamético, ou seja, tomando um número igual de indivíduos por progênie ou sementes por matriz $\sigma_n^2 = 0$ e a expressão (15) torna-se

$$N_e = \frac{2f \bar{n}}{4\bar{n}} = 0,5f. \quad (17)$$

Para um grande número médio (\bar{n}) de indivíduos por progênie, a expressão (16) se aproxima da (17).

No sistema misto de reprodução, considerando o controle gamético, ou seja, igual representatividade das matrizes em termos de números de descendentes, o tamanho efetivo (derivado da equação 13, através da consideração de $\sigma_n^2 = 0$) é dado por:

$$N_e = \frac{2(2-S) f \bar{n}}{(1+S)^2 \bar{n} + (3-2S-S^2)} \quad (18)$$

Esta última expressão é idêntica à apresentada por Vencovsky & Crossa (1999) no contexto da coleta e conservação de germoplasma de populações com sistema reprodutivo misto.

Desconsiderando a estrutura de famílias, uma expressão geral para o tamanho efetivo em populações com sistema misto de reprodução foi apresentada por Gale & Lawrence (1984):

$$N_e = N(1-0,5S) \quad (19)$$

em que N refere-se ao número total de indivíduos ou sementes amostradas e S refere-se à taxa de autofecundação.

A expressão (19) equivale exatamente à expressão (4) e também à expressão (18) considerando a amostragem de apenas um indivíduo por matriz ($\bar{n} = 1$ no denominador da expressão 18). Constatase, assim, que as fórmulas (para espécies alógamas, autógamas e com sistema reprodutivo misto) descritas na literatura sobre o assunto são casos particulares da expressão geral apresentada em (12).

Capítulo 19

**Espécies de Reprodução
Vegetativa e Testes Clonais**

19.1. Introdução

A maioria das espécies perenes e semiperenes permite a propagação vegetativa. O uso da propagação vegetativa na fixação de genótipos superiores, com vistas aos plantios clonais dessas espécies, já foi abordado nos demais capítulos, que enfatizaram a predição de valores genotípicos a partir de testes de progênies ou testes clonais.

Entretanto, em muitas destas espécies, os testes de progênies já são estabelecidos com indivíduos repetidos por meio de seus ramegos, de forma que se tem, simultaneamente, em um único experimento, um teste clonal e um teste de progênies. A utilidade deste tipo de teste tem sido há muito reconhecida (Burdon & Shelbourne, 1974), mas seu uso prático como um fator de aumento da eficiência seletiva é mais recente (Foster & Shaw, 1988). Resende & Araújo (1993) apresentaram procedimentos biométricos para estimação dos componentes de variância genética aditiva e de dominância, herdabilidades no sentido restrito e amplo, predição de valores genéticos aditivos e genotípicos associados à avaliação de famílias de meios-irmãos e de irmãos germanos, com indivíduos repetidos por meio de seus ramegos. As grandes vantagens deste tipo de teste são: (i) permitir a estimativa da variância aditiva e de dominância e herdabilidades no sentido restrito e amplo, (ii) maximizar a eficiência seletiva em um teste de progênies. O fator (ii) é muito relevante, podendo apresentar superioridade em relação à seleção auxiliada por marcadores (MAS), à seleção com S₁'s e à seleção baseada em cruzamentos fatoriais.

19.2. Modelos Estatísticos Lineares

Os procedimentos descritos a seguir baseiam-se em Resende & Araujo (1993).

Modelo genético

Considerou-se o modelo $g = \mu + a + d$, em que g é o efeito genotípico, a é o efeito aditivo e d , o desvio associado à dominância alélica. Portanto, neste modelo que ignora os efeitos epistáticos, a variância genética total (σ_g^2) entre indivíduos de uma população pode ser expressa: $\sigma_g^2 = \sigma_a^2 + \sigma_d^2$, em que: σ_a^2 : variância genética aditiva; σ_d^2 : variância genética de dominância.

Modelo linear geral

Considere a avaliação de indivíduos repetidos por meio de clonagem, diferentes famílias e diferentes procedências ou populações. Para obter o máximo possível de informações, bem como precisão na avaliação dos indivíduos, recomenda-se a distribuição dos propágulos de cada indivíduo em várias parcelas (repetições) com uma ou várias plantas. É desejável também a adoção de um esquema hierárquico de indivíduos dentro de família e família dentro de proce-

dências, segundo o delineamento em blocos de famílias compactas. Nesta situação, empregando-se o delineamento em blocos casualizados, o seguinte modelo linear pode ser especificado:

$$Y_{ijklm} = \mu + p_i + f_{j(i)} + b_k + pb_{ik} + fb_{j(i)k} + c_{l(j(i))} + cb_{l(j(i))k} + e_{ijklm}, \quad (1), \text{ em que:}$$

Y_{ijklm} : observação referente ao propágulo m, do clone l, da família j, da procedência i, no bloco k;

μ : média geral, fixa, $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$;

p_i : efeito da procedência i, $i = 1, 2, \dots, p$, aleatório, $E(p_i) = 0$ e $E(p_i^2) = \sigma_p^2$.

$f_{j(i)}$: efeito da progénie j dentro da procedência i, $j = 1, 2, \dots, f$, aleatório, $E(f_{j(i)}) = 0$ e $E(f_{j(i)}^2) = \sigma_{(F/P)}^2$.

b_k : efeito do bloco k, $k = 1, 2, \dots, b$, fixo, $E(b_k) = b_k$ e $E(b_k^2) = V_b$.

pb_{ik} : interação da procedência i com o bloco k, aleatório, $E(pb_{ik}) = 0$ e $E(pb_{ik}^2) = \sigma_{PB}^2$.

$fb_{j(i)k}$: interação da família j, dentro da procedência i, com o bloco k, aleatório, $E(fb_{j(i)k}) = 0$ e $E(fb_{j(i)k}^2) = \sigma_{(F/P)xB}^2$.

$c_{l(j(i))}$: efeito do clone l, dentro da família j, dentro da procedência i, $l = 1, 2, \dots, c$, aleatório, $E(c_{l(j(i))}) = 0$ e $E(c_{l(j(i))}^2) = \sigma_{C/F/P}^2$.

$cb_{l(j(i))k}$: efeito da interação do clone lji com o bloco k, aleatório, $E(cb_{l(j(i))k}) = 0$ e $E(cb_{l(j(i))k}^2) = \sigma_{(C/F/P)xB}^2$.

e_{ijklm} : efeito do propágulo m do clone l da família j da procedência i no bloco k, $m = 1, 2, \dots, n$, aleatório, $E(e_{ijklm}) = 0$ e $E(e_{ijklm}^2) = \sigma_e^2$.

A análise dos dados obtidos a partir deste delineamento pode ser realizada em três etapas: inicialmente, em termos de famílias dentro de procedências, depois, no âmbito de procedências e, por fim, uma análise global segundo o modelo (1).

Modelos lineares para análise de famílias dentro de cada procedência isolada

a) Experimentos com várias plantas por parcela

$$Y_{ijklm} = \mu + f_j + b_k + fb_{jk} + c_{l(j)} + cb_{l(j)k} + e_{ijklm} \quad (2)$$

b) Experimentos com uma planta por parcela

$$Y_{jkl} = \mu + f_j + b_k + fb_{jk} + c_{l(j)} + cb_{l(j)k} \quad (3)$$

Modelo linear para análise de procedência

$$Y_{jk} = \mu + p_i + b_k + pb_{ik} \quad (4)$$

Modelo linear para análise global de experimentos sem indivíduos repetidos por clonagem

$$Y_{ijklm} = \mu + p_i + f_{j(i)} + b_k + pb_{ik} + fb_{j(i)k} + e_{ijkl} \quad (5)$$

19.3. Análise de Variância, Esperança de Quadrados Médios e Testes dos Efeitos

Nas Tabelas 19.1, 19.2, 19.3, 19.4 e 19.5 são apresentadas as esperanças de quadrados médios associados aos modelos (1), (2), (3), (4) e (5), respectivamente.

Tabela 19.1. Esperança matemática dos quadrados médios [E(QM)] associados ao modelo 1. Teste de procedência/progênie com indivíduos repetidos por clonagem.

F.V.	QM	E(QM)	F
Blocos	Q ₁	-	-
Procedências (P)	Q ₂	$\sigma_E^2 + n\sigma_{(C/F/P)xB}^2 + nc\sigma_{(F/P)xB}^2 + nb\sigma_{C/F/P}^2 + ncbo_{F/P}^2 + ncfb\sigma_{PB}^2$	$\frac{Q_2 + Q_7}{Q_3 + Q_4}$
B x P	Q ₃	$\sigma_E^2 + n\sigma_{(C/F/P)xB}^2 + nc\sigma_{(F/P)xB}^2 + ncfb\sigma_{FB}^2$	Q_3 / Q_7
Famílias(F)/P	Q ₄	$\sigma_E^2 + n\sigma_{(C/F/P)xB}^2 + nc\sigma_{(F/P)xB}^2 + nb\sigma_{C/F/P}^2 + ncbo_{F/P}^2$	$\frac{Q_4 + Q_7}{Q_5 + Q_6}$
(F/P) x B	Q ₆	$\sigma_E^2 + n\sigma_{(C/F/P)xB}^2 + nc\sigma_{(F/P)xB}^2$	Q_7 / Q_8
Clones/F/P	Q ₅	$\sigma_E^2 + n\sigma_{(C/F/P)xB}^2 + nb\sigma_{C/F/P}^2$	Q_5 / Q_7
(C/F/P) x B	Q ₇	$\sigma_E^2 + n\sigma_{(C/F/P)xB}^2$	Q_7 / Q_8
Dentro	Q ₈	σ_E^2	-

Tabela 19.2. Esperança matemática dos quadrados médios [E(QM)] associados ao modelo 2. Teste de progênie com indivíduos repetidos por clonagem - parceiras com várias plantas.

FV	QM	E(QM)	F
Blocos (B)	Q ₁	-	-
Famílias (F)	Q ₂	$\sigma_E^2 + n\sigma_{(C/F)xB}^2 + nc\sigma_{FB}^2 + nb\sigma_{C/F}^2 + ncbo_F^2$	$\frac{Q_2 + Q_5}{Q_3 + Q_4}$
F x B	Q ₄	$\sigma_E^2 + n\sigma_{(C/F)xB}^2 + nc\sigma_{FB}^2$	Q_4 / Q_5
Clones/F	Q ₃	$\sigma_E^2 + n\sigma_{(C/F)xB}^2 + nb\sigma_{C/F}^2$	Q_3 / Q_5
(C/F) x B	Q ₅	$\sigma_E^2 + n\sigma_{(C/F)xB}^2$	Q_5 / Q_6
Dentro	Q ₆	σ_E^2	-

Tabela 19.3. Esperança matemática dos quadrados médios [E(QM)] associados ao modelo 3. Teste de progênies com indivíduos repetidos por clonagem - parcelas com uma planta.

FV	QM	E(QM)	F
Blocos (B)	Q ₁	-	-
Famílias (F)P	Q ₂	$\sigma_{(C/F)xB}^2 + c\sigma_{FB}^2 + b\sigma_{C/F}^2 + cb\sigma_F^2$	$\frac{Q_2 + Q_5}{Q_3 + Q_4}$
Fx B	Q ₄	$\sigma_{(C/F)xB}^2 + c\sigma_{FB}^2$	
Clones/F	Q ₃	$\sigma_{(C/F)xB}^2 + b\sigma_{C/F}^2$	Q_3/Q_5
(C/F) x B	Q ₅	$\sigma_{(C/F)xB}^2$	

Tabela 19.4. Esperança matemática dos quadrados médios [E(QM)] associados ao modelo 4. Teste de procedências, análise ao nível de médias de procedências.

FV	QM	E(QM)	F
Blocos (B)	Q ₁	-	-
Procedências (P)	Q ₂	$\sigma_{PB}^2 + b\sigma_P^2$	Q_2/Q_3
B x P	Q ₃	σ_{PB}^2	

Tabela 19.5. Esperança matemática dos quadrados médios [E(QM)] associados ao modelo 5. Teste de procedência/progénie sem indivíduos repetidos por clonagem.

FV	QM	E(QM)	F
Blocos	Q ₁	-	-
Procedências (P)	Q ₂	$\sigma_E^2 + n\sigma_{(F/P)xB}^2 + nb\sigma_{F/P}^2 + nf\sigma_{PB}^2 + nfb\sigma_P^2$	$\frac{Q_2 + Q_5}{Q_3 + Q_4}$
B x P	Q ₃	$\sigma_E^2 + n\sigma_{(F/P)xB}^2 + nf\sigma_{PB}^2$	Q_3/Q_5
Famílias(F)/P	Q ₄	$\sigma_E^2 + n\sigma_{(F/P)xB}^2 + nb\sigma_{F/P}^2$	Q_4/Q_5
(F/P) x B	Q ₅	$\sigma_E^2 + n\sigma_{(F/P)xB}^2$	Q_5/Q_6
Dentro	Q ₆	σ_E^2	-

Quanto ao teste F e aos testes de diferenças entre médias, pode ser afirmado em relação ao efeito de procedências: a) com número igual de progénies/procedência (ou observações por procedência), os testes devem seguir o modelo 4 (Tabela 19.4); b) com número desigual de progénies (ou observações) por procedência, o modelo 4 não deve ser utilizado, uma vez que as médias de

procedências serão estimadas com diferentes precisões, e, portanto, não podem ser comparadas por meio de um resíduo comum. Neste caso, recomenda-se a utilização do modelo 1, em que o teste F para comparação de procedências será inexato (Tabela 19.1). Neste caso, os graus de liberdade do denominador de F devem ser calculados pela fórmula:

$$g.l. = \frac{(Q_3 + Q_4 - Q_7)^2}{\frac{(Q_3)^2}{g.l._3} + \frac{(Q_4)^2}{g.l._4} + \frac{(Q_7)^2}{g.l._7}}$$

de Satterthwaite (1946), conforme adaptado de Tietjen & Moore (1968).

Há que se ressaltar que, com número desbalanceado de progênies por procedência, o coeficiente do componente de variância σ_{pb}^2 (Tabelas 19.1 e 19.5) diferirá para as fontes procedências e bloco x procedência (Torggler, 1987) e, portanto, o teste F pela fórmula apresentada é apenas aproximado.

Uma opção para teste exato da significância do efeito procedência surge quando a fonte de variação bloco x procedência é não significativa ou negligível. Assim, o efeito procedência pode ser testado diretamente contra o quadrado médio da fonte de variação progénie/procedência.

Nestes casos de desbalanceamento, a comparação entre médias de procedências é imprecisa. Uma alternativa seria a realização de repetidas análises, tomando-se grupos aleatórios de progênies por procedência e, posterior, ponderação das estimativas obtidas de cada grupo. O número de famílias em cada grupo seria, então, aquele associado à procedência com o menor número. Conforme o exposto, pode-se concluir que a melhor solução é evitar o desbalanceamento, por ocasião da implantação do experimento. Entretanto, isto nem sempre é possível do ponto de vista prático. Em geral, nos casos de desbalanceamento, análises mais precisas podem ser efetuadas empregando-se a metodologia de modelos mistos (ver capítulos 7 e 9). Os comentários realizados são válidos para situações de testes de procedência/progénie, de maneira geral, e não apenas para casos em que os indivíduos são repetidos por clonagem.

19.4. Estimação dos Componentes da Variância Genética e Parâmetros Genéticos

Com base no modelo genético considerado, as seguintes igualdades são:

$\sigma_E^2 = \sigma_E^2 = \sigma_{ad}^2$: variância ambiental entre plantas dentro de parcela.

$\sigma_{C/F/P}^2 = \sigma_{C/F}^2$: variância genética entre clones dentro de famílias, equivalente a $(1-\rho_d) \sigma_a^2 + (1-\rho_d) \sigma_d^2$, em que ρ_a é igual ao dobro do coeficiente de parentesco de Malecot e ρ_d é o coeficiente de parentesco de dominância (Kempthorne, 1957). O valor de ρ_a equivale a 1/4 para

meios-irmãos e 1/2 para irmãos germanos, enquanto ρ_d equivale a 0 para meios-irmãos e 1/4 para irmãos germanos;

$\sigma_{F/P}^2 = \sigma_F^2$: variância genética entre famílias, equivalendo a $\rho_a \sigma_a^2 + \rho_d \sigma_d^2$.

σ_p^2 : variância genética entre procedências, a qual inclui efeitos aditivos e não aditivos.

As variâncias genética aditiva (σ_a^2) e de dominância (σ_d^2) podem assim ser estimadas, no caso de famílias de meios irmãos,

$$\sigma_a^2 = 4\sigma_F^2$$

$$\sigma_d^2 = \sigma_{C/F}^2 - 3\sigma_F^2$$

e de irmãos germanos

$$\sigma_a^2 = 3\sigma_F^2 - \sigma_{C/F}^2$$

$$\sigma_d^2 = 2(\sigma_{C/F}^2 - \sigma_F^2)$$

De posse dessas estimativas, os seguintes coeficientes de herdabilidade podem ser definidos:

Modelo 1.

(a) Associados à propagação clonal - válidos para famílias de meios-irmãos e irmãos germanos.

$h_p^2 = \sigma_p^2 / \left(\sigma_p^2 + \frac{\sigma_{PB}^2}{b} + \frac{\sigma_{F/P}^2}{f} + \frac{\sigma_{C/F/P}^2}{fc} + \frac{\sigma_{(F/P)xB}^2}{fb} + \frac{\sigma_{(C/F/P)xB}^2}{cfb} + \frac{\sigma_E^2}{ncfb} \right)$: herdabilidade em termos de médias de procedências.

$h_f^2 = \sigma_{F/P}^2 / \left(\sigma_{F/P}^2 + \frac{\sigma_{C/F/P}^2}{c} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_{(C/F/P)xB}^2}{cb} + \frac{\sigma_E^2}{ncb} \right)$: herdabilidade referente a médias de famílias.

$h_{mif}^2 = \sigma_{C/F/P}^2 / \left(\sigma_{C/F/P}^2 + \frac{\sigma_{(C/F/P)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$: herdabilidade das médias de indivíduo dentro de famílias.

$h_{mi}^2 = (\sigma_p^2 + \sigma_{F/P}^2 + \sigma_{C/F/P}^2) / \left(\sigma_p^2 + \sigma_{F/P}^2 + \sigma_{C/F/P}^2 + \frac{\sigma_{PB}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_{(C/F/P)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$: herdabilidade das médias de indivíduo, considerando o componente de variância entre blocos como negligenciável.

(b) Associados à propagação sexuada - progénies de meios-irmãos.

$h_p^2 =$ idem item anterior (a), porém aplicável somente à seleção de procedências individuais (ausência de cruzamento entre procedências selecionadas).

$h_f^2 =$ idem item anterior (a).

$$h_{mif}^2 = 3\sigma_{F/P}^2 \left(\sigma_{C/F/P}^2 + \frac{\sigma_{(C/F/P)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$$h_{mi}^2 = 4\sigma_{F/P}^2 \left(\sigma_p^2 + \sigma_{F/P}^2 + \sigma_{C/F/P}^2 + \frac{\sigma_{PB}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_{(C/F/P)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

(c) Associados à propagação sexuada - progênies de irmãos germanos.

h_p^2 = idem item anterior (b).

$$h_f^2 = [(3\sigma_{F/P}^2 - \sigma_{C/F/P}^2)/2] \left(\sigma_{FP}^2 + \frac{\sigma_{C/F/P}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_{(C/F/P)xB}^2}{cb} + \frac{\sigma_E^2}{ncb} \right)$$

$$h_{mif}^2 = [(3\sigma_{F/P}^2 - \sigma_{C/F/P}^2)/2] \left(\sigma_{C/F/P}^2 + \frac{\sigma_{(C/F/P)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$$h_{mi}^2 = [(3\sigma_{F/P}^2 - \sigma_{C/F/P}^2)/] \left(\sigma_p^2 + \sigma_{F/P}^2 + \sigma_{C/F/P}^2 + \frac{\sigma_{PB}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_{(C/F/P)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

Nas herdabilidades das médias de indivíduo, associadas à propagação sexuada (itens b e c), desconsiderou-se, nos numeradores das expressões, a variância genética aditiva presente na variação genética entre procedências.

Modelo 2.

(a) Associados à propagação clonal - válidos para famílias de meios-irmãos e irmãos germanos.

$$h_f^2 = \sigma_F^2 / \left(\sigma_F^2 + \frac{\sigma_{C/F}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_{(C/F)xB}^2}{cb} + \frac{\sigma_E^2}{ncb} \right)$$

$$h_{mif}^2 = \sigma_{C/F}^2 / \left(\sigma_{C/F}^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$$h_{mi}^2 = (\sigma_{C/F}^2 + \sigma_F^2) / \left(\sigma_{C/F}^2 + \sigma_F^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$h_{gi}^2 = (\sigma_{C/F}^2 + \sigma_F^2) / (\sigma_{C/F}^2 + \sigma_F^2 + \sigma_{(C/F)xB}^2 + \sigma_{FB}^2 + \sigma_E^2)$: herdabilidade no sentido amplo em termos de indivíduo.

(b) Associados à propagação sexuada - progênies de meios-irmãos.

h_f^2 = idem item (a) do Modelo 2.

$$h_{mif}^2 = 3\sigma_F^2 / \left(\sigma_{C/F}^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$$h_{mi}^2 = 4\sigma_F^2 / \left(\sigma_{C/F}^2 + \sigma_F^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$h_n^2 = 4\sigma_F^2 / (\sigma_{C/F}^2 + \sigma_F^2 + \sigma_{(C/F)xB}^2 + \sigma_{FB}^2 + \sigma_E^2)$: herdabilidade individual no sentido restrito.

(c) Associados à propagação sexuada - progênieis de irmãos germanos.

$$h_f^2 = [(3\sigma_{F/P}^2 - \sigma_{C/F}^2)/2] / \left(\sigma_F^2 + \frac{\sigma_{C/F}^2}{c} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_{(C/F)xB}^2}{cb} + \frac{\sigma_E^2}{ncb} \right)$$

$$h_{mif}^2 = [(3\sigma_F^2 - \sigma_{C/F}^2)/2] / \left(\sigma_{C/F}^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$$h_{mi}^2 = (3\sigma_F^2 - \sigma_{C/F}^2) / \left(\sigma_{C/F}^2 + \sigma_F^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_{FB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$

$$h_n^2 = (3\sigma_F^2 - \sigma_{C/F}^2) / (\sigma_{C/F}^2 + \sigma_F^2 + \sigma_{(C/F)xB}^2 + \sigma_{FB}^2 + \sigma_E^2)$$

Modelo 3

Expressões idênticas às apresentadas para o modelo 2, porém desconsiderando-se o componente σ_E^2 .

Modelo 4

$$h_p^2 = \sigma_p^2 / \left(\sigma_p^2 + \frac{\sigma_{PB}^2}{b} \right)$$

Modelo 5

$\hat{\sigma}_p^2 = [(Q_2 + Q_5) - (Q_3 + Q_4)]/(nfb)$: estimador da variância entre procedências.

$\hat{\sigma}_{F/P}^2 = (Q_4 - Q_5)/nb$: estimador da variância entre progênieis.

$\hat{\sigma}_{PB}^2 = (Q_3 - Q_5)/nf$: estimador da variância da interação procedência x bloco.

$\hat{\sigma}_{(F/P)xB}^2 = (Q_5 - Q_6)/n$: estimador da variância da interação progênie/procedência x bloco.

$\hat{\sigma}_E^2 = Q_6$: estimador da variância dentro de parcelas.

$$h_p^2 = \sigma_p^2 / \left(\sigma_p^2 + \frac{\sigma_{PB}^2}{b} + \frac{\sigma_{F/P}^2}{f} + \frac{\sigma_{(F/P)xB}^2}{fb} + \frac{\sigma_E^2}{nfb} \right)$$
: herdabilidade das médias de procedências.

$$h_f^2 = \sigma_{F/P}^2 / \left(\sigma_{F/P}^2 + \frac{\sigma_{(F/P)xB}^2}{b} + \frac{\sigma_E^2}{nb} \right)$$
: herdabilidade das médias de famílias.

$h_d^2 = (4\sigma_{F/P}^2) / (\sigma_p^2 + \sigma_{F/P}^2 + \sigma_{PB}^2 + \sigma_{(F/P)xB}^2 + \sigma_E^2)$: herdabilidade individual no sentido restrito.

19.5. Métodos de Seleção

Além das modalidades de seleção associadas aos coeficientes de herdabilidade derivados no item anterior, outras mais eficientes, baseadas em índices de seleção, podem ser empregadas.

Estimação de valores genotípicos (V_{G_p})

$$V_{G_p} = h_{mif}^2(MI - MF) + h_f^2(MF - MP) + h_p^2(MP - MG) \text{ (fórmula aproximada)}$$

em que:

MI : média das observações de determinado indivíduo ou clone.

MF : média das observações de determinada família.

MP : média das observações de determinada procedência.

MG : média geral do experimento.

h_{mif}^2 , h_f^2 , h_p^2 : estimadores definidos no item (a) de cada modelo.

Estimação de valores genéticos aditivos (V_G)

$$V_G = h_{mif}^2(MI - MF) + h_f^2(MF - MP) \text{ (fórmula aproximada)}$$

em que:

h_{mif}^2 , h_f^2 : estimadores definidos nos itens (b) e (c) dos modelos 1, 2 e 3.

A partir do delineamento experimental apresentado é possível melhorar a eficiência da seleção, por meio da seleção individual no âmbito das médias de observações, fato este que permite a seleção de indivíduos já testados (conforme conceito de Burdon, 1992).

A metodologia é extremamente útil no estudo de níveis de variabilidade genética para caracteres associados ao enraizamento de estacas e cultivo in vitro. Vários estudos têm sido realizados neste sentido, relatando a presença de variabilidade entre indivíduos (Wilson, 1992; Doran et al., 1992), entre famílias (Wilson, 1992; Willyams et al., 1992a e b) e entre procedências (Willyams et al., 1992a). Porém, esta distribuição dos níveis de variabilidade não tem sido quantificada. Os modelos apresentados, adequados aos casos em que existem repetições clonais de indivíduos, situação predominante em estudos biológicos in vitro, são, portanto, de grande utilidade e devem ser utilizados.

19.6. Seleção a Partir de Um Experimento ou Fase

Para espécies perenes e semiperenes em que se objetiva a propagação assexuada dos indivíduos selecionados ou a propagação sexuada, mas com a recombinação dos próprios indivíduos avaliados, os índices de seleção podem ser melhor explicitados conforme realizado a seguir.

Com a implantação do experimento conforme mostrado anteriormente, o modelo (2) linear associado pode ser assim reescrito:

$$Y_{jklm} = \mu + (\bar{y}_{j...} - \bar{y}_{....}) + (\bar{y}_{..k.} - \bar{y}_{....}) + (\bar{y}_{j..k.} - \bar{y}_{j...} - \bar{y}_{..k.} + \bar{y}_{....}) +$$

$$(\bar{y}_{j..l.} - \bar{y}_{j...}) + (\bar{y}_{jkl.} - \bar{y}_{j..l.} - \bar{y}_{j..k.} + \bar{y}_{j...}) + (y_{jklm} - \bar{y}_{jkl.})$$

Com indivíduos repetidos por reprodução clonal, os efeitos $f_j = (\bar{y}_{j...} - \bar{y}_{....})$ e $c_{\ell(j)} = (\bar{y}_{j.\ell} - \bar{y}_{j...})$ afetam o mérito genético aditivo e genotípico. Assim os índices são:

$I_1 = b_1 (\bar{y}_{j.\ell} - \bar{y}_{j...}) + b_2 (\bar{y}_{j...} - \bar{y}_{....})$, para a predição dos efeitos genéticos aditivos (propagação sexuada) e

$I_2 = b_3 (\bar{y}_{j.\ell} - \bar{y}_{j...}) + b_4 (\bar{y}_{j...} - \bar{y}_{....})$, para a predição dos efeitos genotípicos (propagação assexuada), em que:

$$b_1 = \frac{(1 - \rho_a) \sigma_a^2}{\sigma_{C/F}^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_e^2}{nb}};$$

$$b_2 = \frac{\left[\frac{1 + (c-1) \rho_a}{c} \right] \sigma_a^2}{\sigma_f^2 + \frac{\sigma_{(C/F)}^2}{c} + \frac{\sigma_{fb}^2}{b} + \frac{\sigma_{(C/F)xB}^2}{cb} + \frac{\sigma_e^2}{ncb}},$$

$$b_3 = \frac{(1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2}{\sigma_{C/F}^2 + \frac{\sigma_{(C/F)xB}^2}{b} + \frac{\sigma_e^2}{nb}};$$

$$b_4 = \frac{\left[\frac{1 + (c-1) \rho_a}{c} \right] \sigma_a^2 + \left[\frac{1 + (c-1) \rho_d}{c} \right] \sigma_d^2}{\sigma_f^2 + \frac{\sigma_{(C/F)}^2}{c} + \frac{\sigma_{fb}^2}{b} + \frac{\sigma_{(C/F)xB}^2}{cb} + \frac{\sigma_e^2}{ncb}}.$$

Com uma planta por parcela, não se tem o componente σ_e^2 e os pesos b_1 a b_4 são dados por:

$$b_1 = \frac{[1 - \rho_a] h_a^2}{[1 + (b-1) h_g^2 - b [\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2)] - c_f^2]/b}$$

$$b_2 = \frac{[1 + (c-1) \rho_a] h_a^2}{[1 + (b-1) h_g^2 + (c-1) b [\rho_a h_a^2 + \mu_g (h_a^2 - h_g^2)] + (c-1) c_f^2]/b}$$

$$b_3 = \frac{(1 - \rho_a) h_a^2 + (1 - \rho_d) (h_g^2 - h_a^2)}{[1 + (b-1) h_g^2 - b [(\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2))] - c_f^2]/b}$$

$$b_4 = \frac{[1+(c-1) \rho_a] h_a^2 + [1+(c-1) \rho_d] (h_g^2 - h_a^2)}{[1+(b-1) h_g^2 + (c-1) b [(\rho_a h_a^2 + \rho_d (h_g^2 - h_a^2))] + (c-1) c_f^2]/b}, \text{ em que:}$$

ρ_a e ρ_d : correlações genéticas aditiva e de dominância entre os indivíduos da família, respectivamente (para meios-irmãos $\rho_a = (1/4)$ e $\rho_d = 0$; para irmãos germanos $\rho_a = (1/2)$ e $\rho_d = (1/4)$).

$$h_g^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_y^2} = \frac{\sigma_f^2 + \sigma_{C/F}^2}{\sigma_f^2 + \sigma_{C/F}^2 + \sigma_{(C/F)xB}^2 + \sigma_{FB}^2} : \text{herdabilidade individual no sentido amplo.}$$

$$h_a^2 = \frac{\sigma_a^2}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_f^2 + \sigma_{C/F}^2 + \sigma_{(C/F)xB}^2 + \sigma_{FB}^2} : \text{herdabilidade individual no sentido restrito.}$$

$$c_f^2 = \frac{\sigma_{FB}^2}{\sigma_F^2} = \frac{\sigma_{FB}^2}{\sigma_f^2 + \sigma_{C/F}^2 + \sigma_{(C/F)xB}^2 + \sigma_{FB}^2} : \text{correlação devida ao ambiente comum da parcela de família.}$$

As herdabilidades relativas à média de clone, ignorando a estrutura de famílias, podem também ser expressas em termos das herdabilidades ao nível de indivíduo:

(i) Propagação sexuada

$$h_{m_r}^2 = \frac{b h_a^2}{[1+(b-1) h_g^2]}$$

(ii) Propagação assexuada

$$h_{m_a}^2 = \frac{b h_g^2}{[1+(b-1) h_g^2]}$$

Estes preditores podem ser utilizados para a comparação de métodos de instalação de experimentos (com ou sem repetições clonais de indivíduos) e de seleção, em termos de eficiência seletiva, com um número total fixo de plantas no experimento. Considerando um caráter com $h_a^2 = 0,20$; $h_g^2 = 0,30$ e $c_f^2 = 0,00$

(indivíduos de uma mesma progénie alocados em parcelas não contíguas) e o mesmo tamanho fixo do experimento (100 progénies com 60 indivíduos por progénie para o esquema sem repetições clonais e, 100 progénies, 20 indivíduos por progénie e 3 clones por indivíduo para o esquema com repetições clonais) tem-se os resultados apresentados na Tabela 19.6.

Tabela 19.6. Acurárias** seletivas associadas à experimentação com (C) e sem (S) repetições clonais de indivíduos, para um número total fixo de plantas no experimento.

Delineamento	Tipo de experimento	Propagação sexuada	Propagação assexuada
Fatorial	S	0,68	0,80
Pares simples	S	0,65	0,73
	C	0,71	0,83
Policruzamento	S	0,57	0,65
	C	0,67	0,80
S_1^*	S	0,71	-
Massal	S	0,45	0,55
	C	0,61	0,75

* obtida com base no Capítulo 17; ** acurárias para propagação sexuada: $r_{aa} = [\alpha_1 b_1 + \alpha_2 b_2]^{1/2}$ e

assexuada: $r_{gg} = [r_{aa}^2 + \alpha_5 b_5 + \alpha_6 b_6]^{1/2}$, em que $b_5 = b_3 - b_1$, $b_6 = b_4 - b_2$,

$$\alpha_1 = \frac{c-1}{c}(1-\rho_a), \alpha_2 = \frac{1+(c-1)}{c}\rho_a, \alpha_5 = \frac{c-1}{c}(1-\rho_d) \text{ e } \alpha_6 = \frac{1+(c-1)}{c}(1-\rho_d).$$

Verifica-se pela Tabela 19.6 que, fixando o tamanho do experimento, é muito vantajoso (para caracteres de baixa herdabilidade) instalar testes de progêneres com indivíduos repetidos por reprodução clonal. Para a propagação sexuada, verificam-se: (i) progêneres de meios-irmãos com cada indivíduo representado por três ramegos conduzem, praticamente, à mesma acurácia e ganho do fatorial sem clonagem; (ii) progêneres de irmãos germanos com cada indivíduo representado por três ramegos proporcionam o mesmo ganho que o ganho potencial com S_1 's; (iii) os delineamentos fatorial desconexo e dialélico parcial circulante, com clonagem de indivíduos (resultados não apresentados) certamente são os esquemas que propiciam o maior ganho genético possível. Para a propagação assexuada, verificam-se: (i) progêneres de meios-irmãos com clonagem equivalem ao fatorial sem clonagem; (ii) seleção genotípica pelo índice I_2 em progêneres de irmãos germanos conduz a eficiência de 1,11 (0,83/0,75) sobre a seleção massal clonal baseada em médias de clones (médias de três ramegos) e 1,14 (0,83/0,73) sobre a seleção genotípica por I_2 , porém sem clonagem; (iii) os delineamentos fatorial desconexo e dialélico parcial circulante, com clonagem de indivíduos, certamente são os esquemas que propiciam o maior ganho genotípico possível.

19.7. Vantagens da Clonagem nos Testes de Progêneres e Comparação com a Seleção Auxiliada por Marcadores (MAS)

Para a faixa de herdabilidade (0,20) considerada, a clonagem mostrou-se vantajosa em todas as situações, com eficiências variando de 1,09 (0,71/0,65) a 1,23 (0,80/0,65), para as quatro situações envolvendo as progêneres de meios-

irmãos e irmãos germanos e as propagações sexuada e assexuada (Tabela 19.6). Dessa forma, o uso destes esquemas deve ser estimulado.

O número de indivíduos por progênies escolhido (20) proporciona 87% da acurácia máxima possível em testes não clonados relativos a progênies de meios-irmãos, e o número de ramegos empregado (três) conduz a uma acurácia de 0,75 com a seleção clonal baseada em teste clonal (Resende, 1995). Estes valores, combinados, propiciam uma acurácia de 0,83 para a seleção de clones em testes de progênies de irmãos germanos. Considerando o mesmo tamanho da população experimental, a instalação de teste clonal com 10 ramegos por clone permitiria a avaliação de 600 clones e conduziria a uma acurácia de 0,90. O teste de progênies de irmãos germanos clonado permite a avaliação de 2.000 clones. Assim, em termos de ganho genético com a seleção dos 10 melhores clones, ter-se-ia $G_{s_{IG}} = k r_{IG} = 2,89 \cdot 0,83 = 2,40$ para a seleção no teste de progênies e $G_{s_C} = k r_{IG} = 2,50 \cdot 0,90 = 2,25$ para a seleção no teste clonal, o que fornece uma eficiência de 1,07 em favor do teste com irmãos germanos clonados. O teste não clonado de progênies de irmãos germanos permitiria a avaliação de 6.000 indivíduos e a seleção dos 10 melhores indivíduos para a clonagem forneceria o ganho $G_{s_{IGNC}} = k r_{IG} = 3,23 \cdot 0,73 = 2,35$.

Para a seleção por valores aditivos (população de melhoramento), a seleção dos 100 melhores indivíduos deverá conduzir a ganhos genéticos similares (em torno de 2,05) pelos testes de progênies de irmãos germanos clonados e não clonados.

Assim, justifica-se o uso do teste clonado de progênies de irmãos germanos, o qual permite: (i) eficiência na seleção clonal e por sementes; (ii) estimativa de h_a^2 e h_g^2 .

Em termos de máxima eficiência seletiva possível, falta comparar os testes de progênies clonados com a seleção auxiliada por marcadores (MAS). Trabalho de simulação neste tema foi realizado por Kerr & Goddard (1997), que compararam, entre outras situações, a seguinte: (a) seleção baseada em teste clonal de 10 indivíduos por família de irmãos germanos com cinco ramegos cada um; (b) seleção baseada em marcadores em 50 indivíduos por família de irmãos germanos, com os 10 indivíduos selecionados, avaliados a campo para a aplicação da MAS (dado de campo + dado dos QTL's marcados); (c) seleção em teste de progênies de irmãos germanos não clonados com a avaliação de 50 indivíduos por família. Os ganhos genéticos acumulados após três gerações de seleção foram: 2,85; 2,59 e 2,43 para (a), (b) e (c), respectivamente, considerando um caráter com $h_g^2 = 0,4$ e $\sigma_d^2 / \sigma_a^2 = 1$ (que fornece $h_a^2 = 0,20$ para populações com nível intermediário de melhoramento). Assim, as eficiências do teste clonado sobre a MAS e sobre o teste não clonado foram de 1,10 (2,85/2,59) e 1,17 (2,85/2,43), respectivamente. Por outro lado, a eficiência da MAS sobre o teste não clonado foi de 1,06 (2,59/2,43), resultado este que concorda com as inferências realizadas por Fairfull & Muir (1997).

De maneira geral, Kerr & Goddard (1997) concluíram que os testes clonais conduzem a ganhos comparáveis (iguais ou maiores) àqueles propiciados pela MAS. O uso de QTL's marcados na MAS, empregando o procedimento BLUP, encontra-se descrito no capítulo 7.

Outra vantagem da clonagem dos testes de progênies é a possibilidade de seleção simultânea para a capacidade de enraizamento e produtividade. No caso de espécies do gênero *Eucalyptus*, o próprio crescimento é dependente de uma formação adequada do sistema radicular que, por sua vez, é função do sistema de propagação. Em algumas condições ambientais e para determinadas espécies, materiais propagados por sementes e por micropopulaçao têm apresentado melhor sistema radicular, maior índice de área foliar e também maior crescimento no campo, do que materiais propagados por estaca (MacRae & Cotterill, 1997). Dessa forma, a avaliação dos materiais genéticos propagados já no sistema de propagação a ser utilizado comercialmente, permitirá a seleção daqueles indivíduos que efetivamente crescerão mais nos plantios comerciais.

Ressalta-se que, na escolha da estratégia de propagação e teste, deve-se considerar os diferentes custos associados, e que, provavelmente, as estratégias mais onerosas são, em ordem decrescente de custo: MAS, testes de progênies clonados, testes de progênies não clonados. É importante frisar que, na opção por maior investimento em eficiência de seleção, a clonagem dos testes de progênies empregando-se maior número de ramegos por clone é uma opção que precisa ser considerada, não se justificando o uso indiscriminado apenas das técnicas moleculares.

19.8. Seleção a Partir de Vários Experimentos Seqüenciais

No melhoramento de várias espécies de reprodução vegetativa, os testes clonais são realizados em várias fases, com eliminação gradativa dos clones até a recomendação dos melhores para plantios comerciais. Este é o caso, por exemplo, no melhoramento da cana-de-açúcar (Matsuoka et al., 1999; Barbosa, 2000) e mandioca (Fukuda, 1999) em que, no início, a quantidade de propágulos por planta (obtidas a partir de hibridações) é pequena, porém o número de repetições vai aumentando com o passar das gerações de multiplicação.

Neste caso, duas modalidades gerais de seleção podem ser enunciadas: (i) seleção clonal sem o uso da informação da família; (ii) seleção clonal com o uso da informação de família. No primeiro caso e tendo avaliações de clones em dois experimentos ou fases, pode-se adotar o seguinte índice para combinação das duas fontes de informação (médias fenotípicas) de cada clone:

$$I = b_1 F_1 + b_2 F_2, \text{ em que:}$$

b_1 e b_2 = pesos dados às informações fenotípicas do clone nas fases 1 (F_1) e 2 (F_2), respectivamente.

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sigma_{F_1}^2 & \sigma_g^2 \\ \sigma_g^2 & \sigma_{F_2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_g^2 \\ \sigma_g^2 \end{bmatrix} = \begin{bmatrix} C^{11} & C^{12} \\ C^{12} & C^{22} \end{bmatrix} \begin{bmatrix} \sigma_g^2 \\ \sigma_g^2 \end{bmatrix}.$$

$$C^{11} = \frac{\sigma_{F_2}^2}{\sigma_{F_1}^2 \sigma_{F_2}^2 - \sigma_g^4}; \quad C^{12} = \frac{-\sigma_g^2}{\sigma_{F_1}^2 \sigma_{F_2}^2 - \sigma_g^4}; \quad C^{22} = \frac{\sigma_{F_1}^2}{\sigma_{F_1}^2 \sigma_{F_2}^2 - \sigma_g^4}.$$

Desenvolvendo-se o sistema matricial, obtém-se:

$$b_1 = C^{11} \sigma_g^2 + C^{12} \sigma_g^2 = \frac{\sigma_{F_2}^2 \sigma_g^2 - \sigma_g^4}{\sigma_{F_1}^2 \sigma_{F_2}^2 - \sigma_g^4};$$

$$b_2 = C^{21} \sigma_g^2 + C^{22} \sigma_g^2 = \frac{\sigma_{F_1}^2 \sigma_g^2 - \sigma_g^4}{\sigma_{F_1}^2 \sigma_{F_2}^2 - \sigma_g^4}.$$

Para experimentos implantados com uma planta por parcela, vários propágulos não individualizados por parcela (cana-de-açúcar) ou para variáveis analisadas em termos de médias por parcela e, também, considerando a sua variância ambiental como a mesma nos dois experimentos, tem-se:

$\sigma_{F_1}^2 = \sigma_g^2 + \sigma_e^2 / r_1$: variância fenotípica entre médias de clone no experimento 1, implantado com r_1 ramegos ou repetições por clone.

$\sigma_{F_2}^2 = \sigma_g^2 + \sigma_e^2 / r_2$: variância fenotípica entre médias de clone no experimento 2, implantado com r_2 ramegos ou repetições por clone.

σ_g^2 : variância genotípica.

σ_e^2 : variância ambiental.

Nestas condições, pelo desenvolvimento (substituição de $\sigma_{F_1}^2$ e $\sigma_{F_2}^2$ por $\sigma_g^2 + \sigma_e^2 / r_1$ e $\sigma_g^2 + \sigma_e^2 / r_2$, respectivamente) de b_1 e b_2 acima, obtém-se:

$$b_1 = \frac{r_1}{r_1 + r_2 + \frac{\sigma_e^2}{\sigma_g^2}} = \frac{r_1}{r_1 + r_2 + \frac{1-h_g^2}{h_g^2}}$$

$$b_2 = \frac{r_2}{r_1 + r_2 + \frac{\sigma_e^2}{\sigma_g^2}} = \frac{r_2}{r_1 + r_2 + \frac{1-h_g^2}{h_g^2}}$$

Dessa forma, o valor esperado da média fenotípica do clone ou o seu valor genotípico predito equivale a:

$$I = \frac{r_1}{r_1 + r_2 + \frac{1-h_g^2}{h_g^2}} F_1 + \frac{r_2}{r_1 + r_2 + \frac{1-h_g^2}{h_g^2}} F_2, \text{ em que:}$$

$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_E^2}$: herdabilidade individual no sentido amplo referente ao caráter avaliado.

Verifica-se, assim, que peso maior será dado à média clonal do experimento com maior número de repetições, ou seja, do experimento com maior acurácia.

Este índice pode ser aplicado, também, na seleção de linhagens de espécies autógamas.

Outros pesos podem ser dados às médias fenotípicas, os quais permitem a seleção do material genético correto, mas não fornecem a predição exata dos valores genotípicos. A primeira opção é atribuir peso 1 à média F_1 e obter o peso para b_2 , ou seja:

$$b_1^* = b_1 / b_1 = 1$$

$$b_2^* = b_2 / b_1 = r_2 / r_1$$

Nesta situação, b_2^* é também dado por $b_2^* = \frac{\sigma_{F_1}}{\sigma_{F_2}} \frac{r_{g_{12}} h_2 - r_{F_2} h_1}{h_1 - r_{F_2} r_{g_{12}} h_2}$, em que

$r_{g_{12}} = r_{F_1} / (h_1 h_2) = 1$ e, portanto, $r_{F_1} = h_1 h_2$. Assim:

$$b_2^* = \frac{\sigma_{F_1}}{\sigma_{F_2}} \frac{h_2 - h_1 h_2 h_1}{h_1 - h_1 h_2 h_2} = \frac{\sigma_{F_1}}{\sigma_{F_2}} \frac{h_2(1-h_1^2)}{h_1(1-h_2^2)} = r_2 / r_1, \text{ em que:}$$

$$h_1^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{r_1}}; \quad h_2^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{r_2}}.$$

Uma segunda opção é atribuir para b_1 e b_2 , valores proporcionais a $(b_1 + b_2)$, da seguinte forma:

$$b_1^{**} = \frac{b_1}{b_1 + b_2} = \frac{r_1}{r_1 + r_2}$$

$$b_2^{**} = \frac{b_2}{b_1 + b_2} = \frac{r_2}{r_1 + r_2}$$

Este último índice, $I^{**} = b_1^{**} F_1 + b_2^{**} F_2$, é equivalente ao índice apresentado por Wricke & Weber (1986), Souza Junior (1995) e Souza Junior & Garcia (1996), o qual pode ser generalizado para qualquer número de experimentos por $I = \sum b_i F_i$, em que $b_i = r_i / \sum r_i$, e vem sendo aplicado ao melhoramento de cana-de-açúcar (Matsuoka et al., 1999). É importante ressaltar que o índice I^{**} é fenotípico e não genotípico, pois em comparação com I , considera $h_g^2 = 1$, o que não é verdadeiro.

A acurácia do índice mais realístico, $I = \frac{r_1}{r_1 + r_2 + \frac{1-h_g^2}{h_g^2}} F_1 + \frac{r_2}{r_1 + r_2 + \frac{1-h_g^2}{h_g^2}} F_2$, é dada por $r_{IG} = \sigma_I / \sigma_g$, em que:

$$\sigma_I = [b_1^2 \sigma_{F_1}^2 + b_2^2 \sigma_{F_2}^2 + 2 b_1 b_2 \sigma_{F_1 F_2}]^{1/2}$$

$$= b_1^2 \sigma_{F_1}^2 + b_2^2 \sigma_{F_2}^2 + 2 b_1 b_2 \sigma_g^2$$

$$= \sigma_g \frac{[r_1(r_1 + \lambda) + r_2(r_2 + \lambda) + 2r_1 r_2]^{1/2}}{r_1 + r_2 + \lambda} : \text{desvio padrão do índice.}$$

$\lambda = \frac{\sigma_e^2}{\sigma_g^2} = \frac{1-h_g^2}{h_g^2}$. Assim, $r_{lg} = \frac{[r_1(r_1 + \lambda) + r_2(r_2 + \lambda) + 2r_1 r_2]^{1/2}}{r_1 + r_2 + \lambda}$ e a acurácia na seleção

usando uma só fase equivale a $r_{lg}^* = \left[\frac{r}{r+\lambda} \right]^{1/2}$. Dessa forma, a eficiência do uso do índice em relação ao uso de dados de apenas uma fase pode ser computada por r_{lg} / r_{lg}^* , para cada valor de h_g^2 .

Uma deficiência de todos os índices apresentados é o não ajuste prévio dos dados para os efeitos ambientais dos diferentes experimentos, antes da predição dos valores genotípicos. Como os experimentos apresentam diferentes genótipos quanto ao número e grau de melhoramento (os experimentos correspondentes às fases iniciais apresentam genótipos com menor nível médio de melhoramento do que os genótipos avaliados em fases posteriores), as médias fenotípicas dos diferentes experimentos são viciadas pelo grau de melhoramento dos genótipos que neles se desenvolvem. Dessa forma, recomenda-se o uso do procedimento BLUP para contornar estes problemas. O BLUP para os efeitos genotípicos, neste caso, é dado por:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I \lambda \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ em que:}$$

$$\lambda = \frac{1-h_g^2}{h_g^2}.$$

I : matriz identidade, de ordem equivalente ao número total de clones em avaliação nas várias fases.

$\hat{\beta}$ e \hat{g} : vetores dos efeitos estimados de bloco através dos experimentos e dos efeitos genotípicos preditos, respectivamente.

X e Z : matrizes de incidência (formadas por 0 e 1) associadas a β e g , respectivamente.

y : vetor das médias de clone por repetição dentro de experimentos.

Este procedimento é vantajoso em relação aos índices porque ajusta os dados para os efeitos de blocos e de experimentos. Ademais, enfatiza os dados dos experimentos com maior número de repetições, à semelhança dos índices, e considera a herdabilidade no sentido amplo associada ao caráter em avaliação.

Entretanto, tais procedimentos mencionados anteriormente não são ainda os melhores. Como os clones provêm de cruzamentos entre genitores identificados, existe uma estrutura de família que deve ser considerada na seleção, por propiciar maior ganho genético, conforme demonstrado no tópico 19.6.

Neste caso, na avaliação de clones em dois experimentos seqüenciais, o índice para seleção por valores genotípicos (propagação assexuada) é dado por:

$$I = b_{11} F_{d_1} + b_{12} F_{d_2} + b_{13} F_{e_1} + b_{14} F_{e_2}, \text{ em que:}$$

b_{11} e b_{12} : coeficientes de ponderação dos efeitos de indivíduo dentro de família, nos experimentos 1 (F_{d_1}) e 2 (F_{d_2}), respectivamente.

b_{13} e b_{14} : coeficientes de ponderação dos efeitos de família, nos experimentos 1 (F_{e_1}) e 2 (F_{e_2}), respectivamente.

$$\begin{bmatrix} b_{11} \\ b_{12} \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}^{-1} \begin{bmatrix} G_{11} \\ G_{22} \end{bmatrix}$$

$$\begin{bmatrix} b_{13} \\ b_{14} \end{bmatrix} = \begin{bmatrix} P_{33} & P_{34} \\ P_{43} & P_{44} \end{bmatrix}^{-1} \begin{bmatrix} G_{33} \\ G_{44} \end{bmatrix}$$

$$P_{11} = (1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2 + \sigma_{[(C/F)xB]_1}^2 / r_1$$

$$P_{12} = P_{21} = (1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2$$

$$P_{22} = (1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2 + \sigma_{[(C/F)xB]_2}^2 / r_2$$

$$P_{33} = \rho_a \sigma_a^2 + \rho_d \sigma_d^2 + \sigma_{(F/P)xB]_1}^2 / r_1 + [(1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2] / c + \sigma_{[(C/F)xB]_1}^2 / (cr_1)$$

$$P_{34} = P_{43} = \rho_a \sigma_a^2 + \rho_d \sigma_d^2$$

$$P_{44} = \rho_a \sigma_a^2 + \rho_d \sigma_d^2 + \sigma_{(F/P)xB]_2}^2 / r_2 + [(1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2] / c + \sigma_{[(C/F)xB]_2}^2 / (cr_2)$$

$$G_{11} = G_{22} = (1 - \rho_a) \sigma_a^2 + (1 - \rho_d) \sigma_d^2$$

$$G_{33} = G_{34} = \frac{1 + (c-1) \rho_a}{c} \sigma_a^2 + \frac{1 + (c-1) \rho_d}{c} \sigma_d^2$$

$$F_{d_1} = (Y_{ijk.} - \bar{Y}_{i.k.})$$

$$F_{d_2} = (Y_{ijk'.} - \bar{Y}_{i.k'})$$

$$F_{e_1} = (\bar{Y}_{i,k.} - \bar{Y}_{.,k.})$$

$$F_{e_2} = (Y_{i,k'} - \bar{Y}_{..k'})$$

$Y_{ijk.}$: média do clone j da família i no experimento k.

$\bar{Y}_{ijk.}$: média do clone j da família i no experimento k'.

$\bar{Y}_{i,k.}$: média da família i no experimento k.

$\bar{Y}_{i,k'}$: média da família i no experimento k'.

$\bar{Y}_{..k.}$: média do experimento k.

$\bar{Y}_{..k'}$: média do experimento k'.

Tal índice pode ser aplicado também na seleção por valores genéticos aditivos (propagação sexuada), bastando desconsiderar os componentes associados a σ_d^2 em G_{11} , G_{22} , G_{33} e G_{44} .

O procedimento ideal, que considera a estrutura da família, as herdabilidades nos sentidos restrito e amplo, a estimativa não viciada dos efeitos de blocos, os vários experimentos e a predição simultânea dos efeitos aditivos (a) e de dominância (d), é o BLUP dado por:

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

\hat{a} e \hat{d} : vetores de efeitos genéticos aditivos e de dominância preditos.

A e D : matrizes de correlação genética aditiva e de dominância entre os indivíduos.

$$\lambda_1 = \frac{1-h_g^2}{h_g^2} = \frac{\sigma_e^2}{\sigma_a^2}; \quad \lambda_2 = \frac{1-h_g^2}{(h_g^2 - h_a^2)} = \frac{\sigma_e^2}{\sigma_d^2}.$$

h_a^2 e h_g^2 : herdabilidades individuais nos sentidos restrito e amplo, respectivamente.

y : vetor de médias dos clones por repetição dentro de experimentos.

Os valores genotípicos dos clones são então dados por $\hat{g} = \hat{a} + \hat{d}$, ou podem ser preditos diretamente por:

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

$$g = A \sigma_a^2 + D \sigma_d^2 = Var(g).$$

σ_e^2 : variância ambiental.

As metodologias descritas neste tópico são prontamente aplicáveis no melhoramento de espécies fruteiras (Bruckner, 1999) como o abacaxizeiro, a banana e a laranjeira, espécies forrageiras como o capim-elefante ou napier (Ferreira & Pereira, 1999), espécies olerícolas como o alho, a batata, a batata-doce e o moranguinho e outras espécies como a cana-de-açúcar (Matsuoka et al., 1999; Barbosa, 2000), a mandioca (Fukuda, 1999) e o chá-da-índia (Resende et al., 1998b). Os procedimentos que envolvem a predição do valor genotípico total e que fazem o uso apenas de σ_g^2 (sem a necessidade de desmembrá-la em σ_a^2 e σ_d^2) podem ser aplicados a todas as espécies de reprodução vegetativa, independentemente da ploidia ou mecanismo de reprodução (apomixia ou propágulos vegetativos).

Em cana-de-açúcar, nos esquemas de melhoramento na Austrália são empregados delineamentos de cruzamento muito desbalanceados devido ao florescimento esparso e à baixa fertilidade do pólen. Além disso, a representação das famílias em termos de número de indivíduos é desbalanceada através dos blocos, locais e anos. Em função disto, a partir de 1996 o procedimento BLUP passou a ser usado como técnica rotineira no melhoramento da cana-de-açúcar (Stringer et al., 1996; Stringer, 2000), revelando eficiência e capacidade de reconhecimento precoce de genitores superiores.

19.9. Seleção de Clones Avaliados em Vários Espaçamentos ou Ambientes

Na avaliação de clones não aparentados, em vários espaçamentos (cada espaçoamento em diferentes blocos, permitindo mais de um bloco para cada espaçoamento), o seguinte modelo linear pode ser especificado.

$$Y_{ijkl} = \mu + c_i + b_{j(k)} + s_{ik} + cs_{ik} + e_{ijk} + \delta_{ijkl}, \text{ em que:}$$

μ : média geral.

c_i : efeito aleatório do clone i, $i = 1, 2, \dots, c$.

$b_{j(k)}$: efeito fixo do bloco j, $j = 1, 2, \dots, b$, dentro do espaçoamento k.

s_k : efeito fixo do espaçoamento k, $k = 1, 2, \dots, s$.

cs_{ik} : efeito aleatório da interação do clone i com o espaçoamento k.

e_{ijk} : efeito aleatório da parcela ijk.

δ_{ijkl} : efeito aleatório do indivíduo l dentro da parcela ijk, $l = 1, 2, \dots, n$.

A partir deste modelo básico, várias análises podem ser realizadas, como: (i) análise de clones dentro de cada espaçoamento; (ii) análise conjunta de clones x espaçoamentos; (iii) análise multivariada tratando avaliações de um mesmo clone em diferentes espaçoamentos como sendo caracteres diferentes avaliados em ambientes diferentes.

As análises dentro de cada espaçamento podem ser realizadas conforme descrito para testes clonais no Capítulo 5, ou seja, conforme o esquema:

Fontes de Variação	QM	E(QM)	F
Blocos (B)	-	-	-
Clones (C)	Q ₁	$\sigma_{\delta}^2 / n + \sigma_e^2 + b\sigma_g^2$	Q ₁ /Q ₂
Resíduo (B x C)	Q ₂	$\sigma_{\delta}^2 / n + \sigma_e^2$	-
Dentro de Parcada	Q ₃	σ_{δ}^2	-

A partir desse esquema, os parâmetros relevantes à seleção de clones podem ser estimados por:

$h_{mc}^2 = 1 - (1/F) = \frac{nbh_g^2}{1 + (n-1)(h_g^2 + c^2) + (b-1)nh_g^2}$: herdabilidade no sentido amplo das médias de clones.

$r_{gg} = [h_{mc}^2]^{1/2}$: acurácia na seleção de clones.

$\hat{g} = \mu + [1 - (1/F)](\bar{Y}_{i,k} - \bar{Y}_{..k})$: valor genotípico predito do clone i no espaçamento k.

$\hat{g} \pm t [(1 - r_{gg}^2) \sigma_g^2]^{1/2}$: intervalo de confiança do valor genotípico predito do clone, em que:

$F = Q_1 / Q_2$: valor de F calculado para a fonte de variação clones.

t : valor tabelado da distribuição t de Student associado a determinado nível de confiança.

$h_g^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2 + \sigma_{\delta}^2)$: herdabilidade individual no sentido amplo .

σ_g^2 : variância genotípica entre clones.

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

A análise conjunta (ao nível de médias) de clones e espaçamentos pode ser realizada conforme o esquema a seguir, considerando o efeito de espaçamento como fixo e o efeito de clone como aleatório:

Fontes de Variação	GL	QM	E(QM)	F
Bloco/Espaçamento	(b-1) s	Q ₁	-	-
Espaçamento (E)	(s-1)	Q ₂	$\sigma_E^2 + b \frac{s}{s-1} \sigma_{cs}^2 + bcV_s$	Q ₂ /Q ₄
Clones (C)	(c-1)	Q ₃	$\sigma_E^2 + bs \sigma_c^2$	$F = Q_3/Q_5$
C x E	(c-1)(s-1)	Q ₄	$\sigma_E^2 + b \frac{s}{s-1} \sigma_{cs}^2$	$F^* = Q_4/Q_5$
Resíduo	(b-1)(c-1) s	Q ₅	σ_E^2	-

A partir deste esquema podem ser estimados os parâmetros:

$\sigma_g^2 = \frac{Q_3 - Q_5}{bs}$: variância entre clones, inflacionada pela interação clone x espaçamento (ambiente).

$\sigma_{sc}^2 = \frac{Q_4 - Q_5}{b} \frac{s-1}{s}$: variância da interação clone x espaçamento.

$\sigma_{c*}^2 = \sigma_c^2 - \sigma_{cs}^2$: variância genotípica entre clones, não inflacionada pela interação clone x espaçamento.

$r_{gg} = \frac{\sigma_{c*}^2}{\sigma_{c*}^2 + \sigma_{cs}^2} \frac{s}{s-1} = \frac{1}{1 + \frac{(F^*-1)s}{(F-1)-(F^*-1)(s-1)}}$: correlação genotípica entre um clone e

ele mesmo em diferentes espaçamentos, considerando homogeneidade de variância genotípica através dos espaçamentos (ambientes).

Uma análise multivariada, que trate de avaliações de um mesmo clone em diferentes espaçamentos como sendo caracteres diferentes, avaliados em ambientes diferentes, permite também a estimação de r_{gg} e a predição multivariada dos valores genotípicos dos clones. Considerando dois espaçamentos, a seleção dos clones deve basear-se nos índices apresentados a seguir, os quais predizem os valores genotípicos:

$I_k = b_{11} (\bar{Y}_{i,k} - \bar{Y}_{..k}) + b_{12} (\bar{Y}_{i,k'} - \bar{Y}_{..k'})$, em que:

I_k : valor genotípico predito para o espaçamento (ambiente) k.

$\bar{Y}_{i,k}$: média do clone i no espaçamento k.

$\bar{Y}_{..k'}$: média geral do espaçamento k'.

$$b_{11} = \frac{h_1^2 - r_{gg}^2 h_1^2 h_2^2}{1 - r_{gg}^2 h_1^2 h_2^2}.$$

$$b_{12} = \frac{(\sigma_{g_1}/\sigma_{g_2}) h_2 (r_{gg} h_2 - r_{gg} h_1^2 h_2)}{1 - r_{gg}^2 h_1^2 h_2^2}.$$

$\sigma_{g_1}^2$: variância genotípica entre clones no espaçamento k = 1.

$\sigma_{g_2}^2$: variância genotípica entre clones no espaçamento k' = 2.

h_1^2 : herdabilidade ao nível de médias de clone no espaçamento k = 1.

h_2^2 : herdabilidade da média de clone no espaçamento k' = 2.

$I_k = b_{21} (\bar{Y}_{i,k} - \bar{Y}_{..k}) + b_{22} (\bar{Y}_{i,k'} - \bar{Y}_{..k'})$, em que:

I_k : valor genotípico predito para o espaçamento (ambiente) k'.

$$b_{21} = \frac{(\sigma_{g_2}/\sigma_{g_1}) h_1 (r_{gg} h_1 - r_{gg} h_2^2 h_1)}{1 - r_{gg}^2 h_1^2 h_2^2}$$

$$b_{22} = \frac{h_2^2 - r_{gg}^2 h_1^2 h_2^2}{1 - r_{gg}^2 h_1^2 h_2^2}$$

De posse dos valores I_k e $I_{k'}$ para os vários clones, estes devem ser ordenados visando à seleção de clones ajustados a espaçamentos específicos, para a recomendação de clones já com o seu espaçamento ideal.

19.10. Interação Genótipo x Ambiente e Estratégias de Avaliação Genética e Seleção

Considerando a avaliação de materiais genéticos em dois ambientes ou dois grupos de ambientes, seis possibilidades de seleção existem, as quais são descritas a seguir:

(a) Seleção direta

Nesta modalidade de seleção, o efeito da interação genótipo x ambiente é capitalizado no progresso genético, uma vez que a seleção é baseada nas observações fenotípicas dos materiais genéticos avaliados no próprio ambiente de plantio comercial. Neste caso, a herdabilidade associada à seleção é:

$$h_x^2 = \frac{\sigma_{g_x}^2}{\sigma_{F_x}^2} : \text{herdabilidade da média de clone (ou progênie) no ambiente } x.$$

$\sigma_{g_x}^2$: variância genotípica entre clones (ou progênies) no ambiente x.

$\sigma_{F_x}^2$: variância fenotípica entre médias de clones (ou progênies) no ambiente x.

Entretanto, o ambiente específico do experimento não equivale exatamente ao ambiente de toda a região de plantio comercial a ser cultivada a partir de material genético selecionado em tal experimento. Em outras palavras, a correlação genética entre sub-regiões dentro de regiões de plantio não equivale a 1, de forma que, alguma interação dentro das zonas de melhoramento permanece, mesmo quando realizado o zoneamento das áreas de experimentação e plantio. Neste caso, para computar o ganho genético deve-se considerar uma herdabilidade dada por $r_{gg_d} h_x^2$, em que r_{gg_d} é a correlação genética entre o desempenho dos materiais genéticos nas sub-regiões dentro da região de plantio.

A seleção direta é o processo que maximiza o ganho genético global observado ao longo de várias regiões de plantio, visto que se baseia na identificação de materiais genéticos de desempenho superior em cada ambiente específico. Os demais processos sempre conduzirão a ganhos iguais ou inferiores a este, a menos que sejam usadas informações de vários ambientes como caracteres auxiliares na seleção para ambientes específicos.

(b) Seleção indireta

Nesta modalidade de seleção, os materiais selecionados em determinado ambiente (região x) são plantados em outra região y. O coeficiente de predição genética associado à seleção equivale a:

$$\Psi_{y/x} = h_y \ r_{gg} \ h_x, \text{ em que:}$$

h_y^2 : herdabilidade ao nível de médias de clones (progênies) no ambiente y.

r_{gg} : correlação genética entre o desempenho do material genético em regiões distintas.

Esta estratégia, embora de menor custo em alguns casos, é a menos eficiente se a interação genótipo x ambiente for de alta magnitude.

Na introdução de materiais genéticos selecionados em outros ambientes e considerando herdabilidades similares, a eficiência seletiva é dada por

$$E_1 = \frac{(b)}{(a)} = \frac{h_y \ r_{gg} \ h_x}{r_{gg_d} \ h_x^2} = \frac{r_{gg}}{r_{gg_d}} = r_{gg}, \text{ quando } r_{gg_d} = 1.$$

De maneira geral, a eficiência da seleção indireta (processo b) sobre a direta (processo a) pode ser comparada fazendo $E_1 = r_{gg} / r_{gg_d}$.

(c) Seleção em um ambiente para plantio neste e em outro ambiente

Neste caso, o coeficiente de predição genética é dado por:

$$\Psi_{(x,y)/x} = \frac{h_y \ r_{gg} \ h_x + h_x^2}{2}$$

Na prática do melhoramento de plantas perenes, esta modalidade é a grande competidora da situação (a), quando se comparam as estratégias de melhoramento para ambientes específicos (processo a) versus melhoramento para todos os ambientes de plantio, com seleção baseada em apenas um ambiente.

Nesta situação, assim como no processo (a), deve-se considerar que, mesmo dentro de uma região, existe uma pequena interação entre sub-regiões de forma que $\Psi_{(x,y)/x}$ passa a ser dada por:

$$\Psi_{(x,y)/x} = \frac{h_y \ r_{gg} \ h_x + r_{gg_d} \ h_x^2}{2}, \text{ em que:}$$

r_{gg} : correlação genética entre o desempenho do material genético nas regiões distintas.

r_{gg_d} : correlação genética entre o desempenho do material genético nas sub-regiões dentro de regiões.

Considerando herdabilidades iguais ($h_x^2 = h_y^2 = h^2$) nos diferentes ambientes, $\Psi_{(x,y)/x} = (1/2)(r_{gg} + r_{gg_d})h^2$. Neste caso, uma comparação entre as estratégias (a) e (c) pode ser dada por:

$$E_2 = \frac{(a)}{(c)} = \frac{r_{gg_d} h^2}{(1/2)[(r_{gg} + r_{gg_d})] h^2} = \frac{r_{gg_d}}{(1/2)(r_{gg} + r_{gg_d})}, \text{ em que:}$$

E_2 : eficiência da estratégia de regionalização do programa de melhoramento (estratégia a) em relação à estratégia (c), ou seja, eficiência da estratégia de utilização de dois locais de experimentação em relação ao uso de apenas um local, visando a ganho genético nos dois locais.

Considerando os resultados apresentados no Capítulo 1, os níveis de interação genótipo x ambiente podem ser classificados em:

Níveis	$(1/2)(r_{gg} + r_{gg_d})$	r_{gg_d}	r_{gg}	E_2
Baixo	0,80	0,85	0,75	1,06
Moderado	0,67	0,77	0,57	1,15
Alto	0,50	0,67	0,33	1,34

Estes resultados mostram que a regionalização é vantajosa quando $(1/2)(r_{gg_d} + r_{gg}) \leq 0,80$, ou seja, quando $r_{gg_d} \leq 0,85$ e $r_{gg} \leq 0,75$. É importante frisar que considerar $h_x^2 = h_y^2$ pode implicar considerar homogeneidade de variância genética ($\sigma_{g_x}^2 = \sigma_{g_y}^2$) e ambiental ($\sigma_{ey}^2 = \sigma_{gy}^2$) (mas não necessariamente).

Considerando $r_{gg_d} = 1$, tem-se $E_2 = \frac{r_{gg_d}}{(1/2)(r_{gg_d} + r_{gg})} = \frac{1}{(1/2)(1+r_{gg})} = \frac{2}{(1+r_{gg})}$. Por

esta expressão, quando $r_{gg} = 0,90, 0,80$ e $0,70$, $E_2 = 1,05, 1,11$ e $1,18$, respectivamente. Genericamente (também para mais do que dois ambientes), $E_2 = [\ell/[1+(\ell-1)r_{gg_m}]]$, em que r_{gg_m} é a correlação genética média envolvendo as combinações do local de experimentação (x) com os demais locais de plantio. Verifica-se que esta eficiência será sempre igual ou superior a 1,0 e mede a eficiência do uso (seleção específica para cada um dos locais) de ℓ locais de experimentação em relação ao uso de apenas um local, visando a ganho genético nos ℓ locais.

(d) Seleção usando outro ambiente como auxiliar

Neste processo, a seleção com vistas ao plantio em um ambiente x é realizada usando a informação do próprio e de outros ambientes. No caso de dois ambientes, o coeficiente de predição genética multivariado é dado por:

$$\Psi_m = h_x^2 \left[1 + \frac{(h_y r_{gg} / h_x - r_{gg} h_y h_x)^2}{1 - r_{gg}^2 h_x^2 h_y^2} \right]^{1/2}$$

A eficiência deste processo seletivo em relação ao descrito em (a) é dada por:

$$E_3 = \frac{(d)}{(a)} = \left[1 + \frac{[r_{gg}(1-h^2)]^2}{1 - r_{gg}^2 h^4} \right]^{1/2}, \text{ considerando } h^2 \text{ similares e } r_{gg_d} = 1. \text{ Esta eficiência será sempre igual ou superior a 1,0.}$$

(e) Seleção baseada na média de dois ambientes para o plantio nos dois ambientes

Esta estratégia é utilizada quando se deseja plantar em um ambiente x, materiais mais estáveis.

Neste caso, a herdabilidade associada à média geral dos materiais genéticos nos dois ambientes é dada por:

$$h_{(x,y)/(x,y)}^2 = \frac{\sigma_g^2 + 1/2 \sigma_{g\ell}^2}{\sigma_g^2 + \sigma_{g\ell}^2 / \ell + \sigma_E^2 / b\ell} = \frac{\sigma_g^2 + 1/2 \sigma_{g\ell}^2}{\left[\frac{1 + (\ell-1) r_{gg} h_{mc}^2}{\ell} \right] \sigma_F^2} = \frac{r_{gg} h_{mc}^2 + \frac{1-r_{gg}}{2} h_{mc}^2}{\frac{1 + (\ell-1) r_{gg} h_{mc}^2}{\ell}} = \frac{(r_{gg} + 1) \ell h_{mc}^2}{2[1 + (\ell-1) r_{gg} h_{mc}^2]}$$

, considerando herdabilidades iguais ($h_{mc}^2 = h_{mcx}^2 = h_{mcy}^2$), em que:

$$h^2 = \frac{\sigma_g^2 + \sigma_{g\ell}^2}{\sigma_{Fm}^2} = \frac{\sigma_g^2 + \sigma_{g\ell}^2}{\sigma_g^2 + \sigma_{g\ell}^2 + \sigma_e^2 / b} : \text{herdabilidade da média de clone em cada ambiente.}$$

σ_{Fm}^2 : variância fenotípica ao nível de médias de clone em cada ambiente.

σ_g^2 : variância genotípica entre clones (progênies) livre da variância ($\sigma_{g\ell}^2$) da interação genótipo x ambiente.

$$r_{gg} = \sigma_g^2 / (\sigma_g^2 + \sigma_{g\ell}^2).$$

Com $\ell = 2$, tem-se:

$$h_{(x,y)/(x,y)}^2 = \frac{(1+r_{gg}) h_{mc}^2}{1+r_{gg} h_{mc}^2} = \frac{1+r_{gg}}{1/h_{mc}^2 + r_{gg}}$$

Esta fórmula pode ser também expressa em termos da herdabilidade (h_i^2) e variância fenotípica individual ($\sigma_{F_i}^2$) (Resende, 1998):

$$\begin{aligned}
h_{(x,y)/(x,y)}^2 &= \frac{\sigma_g^2 + (1/2) \sigma_{g\ell}^2}{1 + (b-1) h_i^2 + b(\ell-1) r_{gg} h_i^2} = \frac{\sigma_g^2 + (1/2) \sigma_{g\ell}^2}{\sigma_{F_i}^2 \left[r_{gg} h_i^2 + \frac{(1-r_{gg}) h_i^2}{\ell} + \frac{1-h_i^2}{b\ell} \right]} = \\
&= \frac{r_{gg} h_i^2 + \frac{1-r_{gg}}{2} h_i^2}{\left[r_{gg} h_i^2 + \frac{1-r_{gg}}{\ell} h_i^2 + \frac{1-h_i^2}{b\ell} \right]}.
\end{aligned}$$

Para dois locais, o ganho genético com esta modalidade (e) de seleção, em termos do diferencial de seleção de um ambiente x, é dado por:

$$\begin{aligned}
G_{s_e} &= k h_{(x,y)/(x,y)}^2 \sigma_{F_{(x,y)}} = k \frac{(1+r_{gg}) h_{mcx}^2}{1+r_{gg} h_{mcx}^2} \left[\frac{1+r_{gg} h_{mcx}^2}{2} \right]^{1/2} \sigma_{F_x} = \\
&= k \sigma_{F_x} \frac{(1+r_{gg}) h_{mcx}^2}{1+r_{gg} h_{mcx}^2} \left[\frac{1+r_{gg} h_{mcx}^2}{2} \right]^{1/2}.
\end{aligned}$$

O ganho genético na modalidade de seleção (c) é dado por:

$$G_{s_c} = k \sigma_{F_x} \frac{h_y r_{gg} h_x + h_x^2}{2}$$

Com $h_x^2 = h_y^2 = h_{mcx}^2$, tem-se:

$$G_{s_c} = k \sigma_{F_x} \frac{(1+r_{gg}) h_{mcx}^2}{2}$$

Dessa forma, a eficiência da modalidade (e) em relação à modalidade (c) é dada por:

$$E_4 = \frac{G_{se}}{G_{sc}} = \frac{\frac{(1+r_{gg}) h_{mcx}^2}{1+r_{gg} h_{mcx}^2} \left[\frac{1+r_{gg} h_{mcx}^2}{2} \right]^{1/2}}{\frac{(1+r_{gg}) h_{mcx}^2}{2}} = \left[\frac{2}{1+r_{gg} h_{mcx}^2} \right]^{1/2}$$

Alternativamente:

$$E_4 = [h_{(x,y)/(x,y)}^2 / \psi_{(x,y)/x}]^{1/2}.$$

Genericamente, para ℓ locais, $E_4 = \left[\frac{\ell}{1 + (\ell-1) r_{gg} h^2} \right]^{1/2}$. Com $h^2 \rightarrow 1$, que significa herdabilidade ao nível de médias em cada local tendendo a 1, tem-se $E_4 = \left[\frac{\ell}{1 + (\ell-1) r_{gg}} \right]^{1/2}$, conforme apresentado por Resende & Rosa-Perez (1999a). E_4 representa a eficiência da seleção baseada na média de ℓ locais de experimentação em relação ao uso de apenas um local, para ganho genético nos ℓ locais.

(f) Seleção baseada na média de dois ambientes para o plantio em um dos ambientes

Este caso é similar ao descrito em (e) e apresenta herdabilidade:

$$h_{x/(x,y)}^2 = \frac{(1+r_{gg}) h^2}{1+r_{gg} h^2} = h_{(x,y)/(x,y)}^2, \text{ quando } h_x^2 = h_y^2 = h_{mcx}^2 = h^2.$$

Este processo pode ser comparado com (a) e com (d) para saber qual é o mais eficiente.

Comparando-se (f) com (a) pode-se computar a eficiência:

$$E_5 = \frac{\frac{(1+r_{gg}) h^2}{1+r_{gg} h^2} \left[\frac{1+r_{gg} h^2}{2} \right]^{1/2}}{h^2} = \frac{(1+r_{gg}) [1+r_{gg} h^2]^{1/2}}{(1+r_{gg} h^2) [2]^{1/2}} = \frac{1+r_{gg}}{[2(1+r_{gg} h^2)]^{1/2}}$$

Comparando-se (d) com (a), pode-se computar a eficiência:

$$E_6 = \left[1 + \frac{(h_y r_{gg}/h_x - r_{gg} h_y h_x)^2}{1 - r_{gg}^2 h_x^2 h_y^2} \right]^{1/2} = \left[1 + \frac{(r_{gg} - r_{gg} h^2)^2}{1 - r_{gg}^2 h^4} \right]^{1/2} = \left[1 + \frac{[r_{gg}(1-h^2)]^2}{1 - r_{gg}^2 h^4} \right]^{1/2} =$$

E_3 , para $h_x^2 = h_y^2$. A eficiência de (f) sobre (d) é dada por E_5/E_6 .

Considerando $r_{gg} = 0,8$ e $h^2 = 0,20$, tem-se $E_5 = 1,1817$ e $E_6 = 1,1918$. Já com $r_{gg} = 0,2$ e $h^2 = 0,2$, tem-se $E_5 = 0,8320$ e $E_6 = 1,0127$. Verifica-se, assim, que, para uma mesma herdabilidade, os dois procedimentos (f e d) conduzem a eficiências similares quando a correlação genética é alta. Quando a correlação é baixa, o procedimento (d) é muito superior ao (f) e similar ao (a).

Com $h^2 = 0,90$ e $r_{gg} = 0,80$, $E_5 = 0,97$ e portanto a estratégia (a), de regionalização do programa leva a uma eficiência de $1/E_5 = 1,03$, em relação à seleção de materiais estáveis (condução de uma única população de melhoramento com seleção baseada em médias de materiais genéticos através dos locais). A eficiência de (a) em relação a (f), pode ser computada pela razão entre acurá-

cias dada por $E_7 = \left[\frac{r_{gg_d} h^2}{(1/2)(r_{gg_d} h^2 + r_{gg} h^2)} \right]^{1/2} = \left[\frac{r_{gg_d}}{(1/2)(r_{gg_d} + r_{gg})} \right]^{1/2}$, quando $h^2 \rightarrow$

1,00 e $r_{gg_d} \neq 1$. Quando $r_{gg_d} = 1$ e $h^2 \rightarrow 1$, $E_7 = \left[\frac{2}{1+r_{gg}} \right]^{1/2}$.

De maneira genérica, quatro situações principais podem ser consideradas na seleção que visa ao plantio em duas regiões ou ambientes:

(i) Seleção em cada ambiente (situação (a))

$$\text{Eficiência: } E_8 = \frac{h^2}{h^2} = 1,0.$$

(ii) Seleção em apenas um ambiente (situação (c))

$$\text{Eficiência em relação a (i): } E_9 = \frac{1 + r_{gg}}{2}.$$

(iii) Seleção baseada na média de dois ambientes (situação (e))

$$\text{Eficiência em relação a (i): } E_{10} = \frac{1 + r_{gg}}{[2(1 + r_{gg} h^2)]^{1/2}}.$$

(iv) Seleção (através de índice) baseada nas informações individualizadas dos dois ambientes (situação (d))

$$\text{Eficiência em relação a (i): } E_{11} = \left[1 + \frac{[r_{gg} (1 - h^2)]^2}{1 - r_{gg}^2 h^4} \right]^{1/2}$$

Considerando $h^2 = 0,90$ (que equivale a uma herdabilidade da média de 20 ramos de um clone ou a uma herdabilidade da média de família com 60 indivíduos, valor este associado a uma herdabilidade individual equivalente a cerca de 0,30), têm-se as seguintes eficiências para diferentes valores de r_{gg} :

r_{gg}	E_8	E_9	E_{10}	E_{11}	E_{10}/E_9
0,90	1,00	0,95	1,00	1,01	1,05
0,80	1,00	0,90	0,97	1,01	1,08
0,70	1,00	0,85	0,94	1,00	1,10
0,50	1,00	0,75	0,88	1,00	1,17
0,30	1,00	0,65	0,82	1,00	1,26

Nesta situação, pode-se concluir:

- E_8 e E_{11} são iguais e, portanto, a seleção em cada ambiente propicia o ganho genético máximo e não é necessário o uso das informações do outro ambiente (isto pode ser explicado pelo alto valor de h^2 considerado).
- A seleção de genótipos mais estáveis (condução de uma única população de melhoramento com seleção pela média) proporciona, praticamente, o mesmo ganho que a regionalização (E_8), quando $r_{gg} \geq 0,80$, conforme pode ser visto por E_{10} ; com $r_{gg} < 0,70$ a regionalização é vantajosa.
- A condução de uma única população de melhoramento com seleção baseada em apenas um ambiente só é satisfatória quando $r_{gg} > 0,90$, conforme pode ser verificado por E_9 e E_{10}/E_9 .

Dessa forma, nestas condições recomendam-se as estratégias:

- (i) Regionalização ou seleção de genótipos específicos: quando $r_{gg} < 0,70$.
- (ii) População de melhoramento única com seleção de materiais estáveis (seleção pela média através dos locais): quando $0,70 \leq r_{gg} < 0,90$.
- (iii) População de melhoramento única com seleção indireta: quando $r_{gg} \geq 0,90$ ou $r_{gg} \geq 0,70$ considerando que a eficiência de $1,10 (E_{10}/E_9)$ implica avaliar o dobro do número de indivíduos.

É importante, na prática, verificar o efeito de ℓ apenas como fator de redução de $\sigma_{g\ell}^2 / \ell$ e não como redutor de σ_g^2 / ℓ e $\sigma_e^2 / n\ell$ em $(\sigma_{g\ell}^2 / \ell + \sigma_E^2 / (n\ell))$.

Os ganhos obtidos quando se faz E_{10}/E_9 implicam avaliar o dobro do número de indivíduos. Assim, é relevante determinar o número adequado de locais de experimentação, considerando um número total fixo de indivíduos a serem avaliados.

O número adequado de locais de experimentação para um número total fixo de indivíduos depende da herdabilidade do caráter e da correlação genética intraclasses através dos locais. Fixando-se em $n\ell$ o número total de indivíduos por acesso, em que n refere-se ao número de indivíduos por local, são encontrados os seguintes estimadores para as variâncias fenotípicas ao nível de médias de acesso, considerando a avaliação dos $n\ell$ indivíduos em um só ambiente ou em vários ambientes (Resende, 1998):

$$\hat{\sigma}_{F\ell}^2 = \left(\hat{h}_i^2 + \frac{1 - \hat{h}_i^2}{n\ell} \right) \hat{\sigma}_{Fi}^2$$

$$\hat{\sigma}_{Fm}^2 = \left(\hat{r}_{gg} \hat{h}_i^2 + \frac{(1 - \hat{r}_{gg})}{\ell} \hat{h}_i^2 + \frac{(1 - \hat{h}_i^2)}{n\ell} \right) \hat{\sigma}_{Fi}^2$$

em que:

$\hat{\sigma}_{F\ell}^2$ e $\hat{\sigma}_{Fm}^2$: estimativas das variâncias fenotípicas das médias de acesso para o local ℓ e na média dos ℓ locais, respectivamente;

$\hat{\sigma}_{Fi}^2$: estimativa da variância fenotípica individual dentro de locais.

$\hat{h}_i^2 = \frac{\hat{\sigma}_g^2 + \hat{\sigma}_{g\ell}^2}{\hat{\sigma}_{Fi}^2}$: estimativa da herdabilidade individual dentro de locais.

A eficiência (E_{12}) da seleção baseada em vários locais em relação à seleção baseada em um só local é dada por:

$$E_{12} = \left[\frac{\hat{\sigma}_g^2 + \hat{\sigma}_{g\ell}^2 + \hat{\sigma}_{\delta}^2 / (n\ell)}{\hat{\sigma}_g^2 + \frac{\hat{\sigma}_{g\ell}^2}{\ell} + \frac{\hat{\sigma}_{\delta}^2}{n\ell}} \right]^{1/2} = \left[\frac{\hat{h}_i^2 + (1 - \hat{h}_i^2) / n\ell}{\hat{r}_{gg} \hat{h}_i^2 + \frac{(1 - \hat{r}_{gg})}{\ell} \hat{h}_i^2 + (1 - \hat{h}_i^2) / n\ell} \right]^{1/2} = \left[\frac{1 + (n\ell - 1) \hat{h}_i^2}{1 + (n - 1) \hat{h}_i^2 + n(\ell - 1) \hat{r}_{gg} \hat{h}_i^2} \right]^{1/2}$$

em que $\hat{\sigma}_{\delta}^2$ refere-se à variação residual dentro de locais.

Na Tabela 19.7, são apresentados valores de eficiência para alguns valores de n , ℓ , e r_{gg} para $h^2 = 0,20$. Para $h^2 = 0,20$, 30 indivíduos por família proporcionam acurárias da ordem de 90% da máxima possível para a seleção de indivíduos para propagação por sementes ou de clones para propagação vegetativa (Resende, 1995).

Tabela 19.7. Eficiência da experimentação em vários locais (ℓ) em relação a um só local, considerando um número total ($n\ell$) fixo de indivíduos avaliados, para várias magnitudes de correlação genética (r_{gg}) intraclasse através dos locais, número (n) de indivíduos por acesso em cada local e herdabilidade (h^2) de 20%.

h^2	$n\ell$	ℓ	n	r_{gg}	E_{12}	Incre- mento
0,20	30	2	15,0	0,30	1,20	20
		3	10,0	0,30	1,30	10
		4	7,5	0,30	1,38	8
		5	6,0	0,30	1,40	2
0,20	30	2	15,0	0,50	1,13	13
		3	10,0	0,50	1,19	6
		4	7,5	0,50	1,22	3
		5	6,0	0,50	1,24	2
0,20	30	2	15,0	0,70	1,07	7
		3	10,0	0,70	1,10	3
		4	7,5	0,70	1,12	2
		5	6,0	0,70	1,13	1

Para um número total fixo de indivíduos avaliados, conclui-se, com base na Tabela 19.7, que se torna vantajosa (ganho 6%) a utilização de 4, 3 e 2 locais para correlações de magnitudes da ordem de 0,30; 0,50 e 0,70, respectivamente.

Na realização de análises sobre interação genótipo x ambiente pela metodologia de modelos mistos, no arquivo de dados para uso do software DFREML (Meyer, 1998) devem constar as seguintes colunas:

Indivíduo Pai Mãe Bloco Ambiente Progênie x Local Parcela Vetor de Dados

Considerando a avaliação de 25 progênies de erva-mate em dois locais, para o caráter peso de folhas, em um total de 18 blocos e com parcelas de 6 plantas, obtiveram-se os seguintes (Tabela 19.8) resultados por três diferentes modelos de análise.

Tabela 19.8. Estimativas de parâmetros genéticos e fenotípicos por três modelos de análise, envolvendo o caráter peso de folhas em erva-mate.

Parâmetros*	Modelo 1: Bloco como efeito fixo	Modelo 2: Blo- co/Ambiente e am- biente como efeitos fixos	Modelo 3: Blo- co/Ambiente como efeito fixo e ambiente como efeito aleatório
h_i^2	0,07	0,07	0,06
σ_a^2	0,013	0,013	0,013
$\sigma_{g\ell}^2$	0,0035	0,0035	0,0035
σ_c^2	0,027	0,027	0,027
σ_ℓ^2	-	-	0,033
σ_F^2	0,1913	0,1913	0,2243
Log L	921,92	921,92	921,92

* h_i^2 : herdabilidade individual; σ_a^2 : variância genética aditiva; $\sigma_{g\ell}^2$: variância da interação genótipo x ambiente; σ_c^2 : variância entre parcelas; σ_ℓ^2 : variância entre locais; Log L: logaritmo da função de verossimilhança maximizada.

Verifica-se, conforme esperado, que os modelos 1 e 2 conduzem a resultados idênticos para as estimativas dos parâmetros genéticos. A única informação adicional que o modelo 2 apresenta é a estimativa dos efeitos de ambiente. Comparando os modelos 1 e 2 com o modelo 3, constata-se que este último, embora apresentando o mesmo Log L, conduz a menor herdabilidade e a maior variância fenotípica, devendo, por isto, ser preferido. Todos os modelos levaram a estimativas idênticas para a variância da interação genótipo x ambiente.

19.11. Análise Ambiental a partir de Experimentos no Delineamento em Blocos

Na análise de experimentos em genética e melhoramento de plantas feitos em um único ambiente, tradicionalmente, tem-se enfatizado mais a análise genética do que a ambiental. Embora a seleção deva basear-se em um ordenamento dos valores genéticos dos indivíduos candidatos à seleção, a utilização prática e a comprovação do valor real dos materiais genéticos melhorados baseiam-se em seus valores fenotípicos (nos plantios comerciais), os quais são influenciados pelo ambiente. Isto justifica uma análise mais detalhada dos efeitos ambientais em um experimento.

A análise ambiental deve enfatizar pelo menos três fatores: (i) a eficiência do delineamento em termos do controle local; (ii) a variabilidade espacial dentro dos estratos ambientais homogêneos (blocos); (iii) a interação genótipo x ambi-

ente dentro de um mesmo sítio ou fazenda. O fator (i) pode ser estudado com base na significância do teste F de Snedecor associado à fonte de variação blocos na análise de variância e também com base no coeficiente de correlação intraclasse entre parcelas dentro dos blocos (ρ_b). A variabilidade espacial dentro dos blocos pode ser estudada por meio do coeficiente de correlação intraclasse entre indivíduos de uma mesma parcela, devido ao ambiente comum da parcela (c^2) o qual pode, alternativamente, ser denominado coeficiente de determinação dos efeitos ambientais entre parcelas. Por sua vez, o fator (iii) pode ser investigado com base na correlação genética intraclasse (ρ_g), dos materiais genéticos ao longo das repetições, ou seja, de uma repetição para outra.

A eficiência do delineamento experimental em blocos (DBC), avaliada a partir de uma análise ao nível de médias de parcela é apresentada a seguir, considerando os efeitos de blocos como aleatórios. Considere os seguintes resultados, associados à avaliação de $c = 62$ clones de caju em $b = 3$ blocos, para o caráter diâmetro da copa.

FV	GL	QM	QM	E (QM)	F
Bloco	2	Q_1	0,41845	$\sigma^2 + c\sigma_b^2 = \sigma^{2*}[1 + (c-1)\rho_b]$	4,14*
Clones	61	Q_2	0,17912	-	-
Resíduo	122	Q_3	0,10106	$\sigma^2 = \sigma^{2*}(1 - \rho_b)$	

* significativo ao nível de 5% de probabilidade de erro tipo I.

Têm-se as seguintes estimativas:

$\hat{\sigma}^2 = Q_3 = 0,10106$: estimativa da variância residual.

$\hat{\sigma}_b^2 = (Q_1 - Q_3)/c = 0,005119$: estimativa da variância entre blocos.

$\hat{\rho}_b = \frac{\hat{\sigma}_b^2}{\hat{\sigma}^2 + \hat{\sigma}_b^2} = 0,04821$: estimativa da correlação intraclasse entre parcelas dentro dos blocos.

Sendo $\hat{\sigma}^{2*} = \hat{\sigma}_b^2 + \hat{\sigma}^2$, a esperança da variância residual para um delineamento inteiramente ao acaso (DIC) e $1/\hat{\sigma}^{2*}$ e $1/\hat{\sigma}^2$ as **quantidades de informação** associadas ao DIC e ao DBC, respectivamente, a eficiência do DBC em relação ao DIC é dada pela razão entre as respectivas quantidades de informação, ou seja, por $E_{13} = (1/\hat{\sigma}^2)/(1/\sigma^{2*}) = \frac{\hat{\sigma}^{2*}}{\hat{\sigma}^2} = \frac{\hat{\sigma}^2 + \hat{\sigma}_b^2}{\hat{\sigma}^2} = \frac{1}{1 - \hat{\rho}_b} = 1,0507$.

Verifica-se, pela expressão de E_{13} , que a eficiência de um DBC em relação a um DIC é tanto maior quanto maior for o coeficiente de correlação intraclasse ρ_b . No presente exemplo, tal eficiência foi baixa, revelando a existência de heterogeneidade entre blocos na área experimental (a estatística F para blocos mostrou-se significativa), mas de baixa magnitude.

A significância dos efeitos de blocos deve ser analisada em conjunto com o parâmetro c^2 . Assim, têm-se quatro situações: (a) F para blocos significativo e c^2 alto; (b) F para blocos significativo e c^2 baixo; (c) F para blocos não significativo e c^2 baixo; (d) F para blocos não significativo e c^2 alto. Um c^2 alto significa alta variabilidade entre parcelas no bloco e um c^2 baixo significa baixa variação entre parcelas no bloco. Considerando o conceito de **capacidade de teste** como a capacidade de o experimento propiciar aos materiais genéticos experimentar diferentes condições ambientais, pode-se fazer as inferências apresentadas a seguir.

Na situação (a), pode-se dizer que o delineamento não foi totalmente eficiente, mas a capacidade de teste foi adequada. Isto porque, embora os blocos tenham apresentado diferenças significativas entre eles, uma grande heterogeneidade ambiental dentro dos blocos permaneceu. Em (b), pode-se inferir que o delineamento foi eficiente e a capacidade de teste adequada. Em (c), existe uma grande homogeneidade ambiental na área experimental e, neste caso, qualquer delineamento é eficiente mas não existe uma capacidade de teste adequada, podendo-se incorrer no risco de selecionar materiais genéticos com pequena **plasticidade fenotípica**. A situação (d), por sua vez, denota que o delineamento não foi eficiente e a capacidade de teste inadequada. Neste último caso, o melhorista deve procurar métodos mais sofisticados de análise, tais como uma análise espacial visando à realização de uma blocagem a posteriori. As causas dos resultados associados à situação (d) podem ser atribuídas a: (i) blocos muito grandes foram alocados, de forma que a variação dentro de blocos tendeu a ser próxima da magnitude da variação entre blocos (a correlação intraclass entre parcelas dentro de bloco, p_b , foi muito baixa); (ii) o gradiente ambiental (de fertilidade, por exemplo) ocorre em vários sentidos. As quatro situações mencionadas encontram-se sintetizadas na Tabela 19.9.

Tabela 19.9. Inferências práticas sobre eficiência de delineamento e capacidade de teste, em função das estatísticas F de Snedecor para blocos e coeficiente de determinação dos efeitos de parcela (c^2).

Situação	F	c^2	Eficiência de delineamento	Capacidade de teste
(a)	Significativo	Alto	Não	Sim
(b)	Significativo	Baixo	Sim	Sim
(c)	Não significativo	Baixo	Sim	Não
(d)	Não significativo	Alto	Não	Não

Em resumo, este tipo de análise deve ser realizado e têm-se as seguintes implicações práticas:

- situação (a): o melhorista deve utilizar métodos mais sofisticados de análise.

- situação (b): é a situação ideal ao melhorista.
- situação (c): o melhorista deve prever maior perda de ganho genético realizado devido à interação genótipo x ambiente.
- situação (d): o melhorista deve lançar mão de métodos mais sofisticados de análise, mas não necessariamente conseguirá uma capacidade de teste adequada (nas situações (a) e (d), o melhorista deveria ter usado outro delineamento, como o látice ou o quadrado latino).

É importante mencionar que os métodos de blocagem a posteriori tenderão a propiciar maiores eficiências de delineamento e capacidades de teste, quando os experimentos forem implantados no delineamento inteiramente casualizado e com uma planta por parcela.

A variabilidade espacial dentro de blocos e a interação genótipo x ambiente dentro de um mesmo sítio podem ser investigadas tomando-se por base uma análise de um DBC ao nível de plantas individuais. Considerando o mesmo experimento com clones de caju, com $n = 4$ plantas por parcela, os resultados da análise de variância são:

FV	GL	QM	QM	E (QM)	F
Bloco (B)	6	Q_1	1,6738	-	-
Clones (C)	61	Q_2	0,71648	$\sigma_\delta^2 + n\sigma_e^2 + nb\sigma_g^2$	$F=1,7724$
Resíduo (BxC)	122	Q_3	0,40424	$\sigma_\delta^2 + n\sigma_e^2$	$F^*=1,1392^*$
Dentro de parcela	558	Q_4	0,35484	σ_δ^2	

* não significativo.

As estimativas dos parâmetros de interesse são:

$\hat{\sigma}_\delta^2 = Q_4 = 0,35484$: estimativa da variância dentro de parcelas.

$\hat{\sigma}_e^2 = (Q_3 - Q_4)/n = 0,01235$: estimativa da variância entre parcelas.

$\hat{\sigma}_g^2 = (Q_2 - Q_3)/nb = 0,02602$: estimativa da variância genotípica entre clones.

$\hat{c}^2 = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_\delta^2 + \hat{\sigma}_e^2 + \hat{\sigma}_g^2} = 0,0314$ estimativa do coeficiente de determinação dos efeitos de parcela.

$\hat{\rho}_g = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2} = \frac{1}{1 + \frac{(F^*-1)}{F-1}} \frac{b}{F^*} = 0,6781$ estimativa da correlação genética intraclasse dos materiais genéticos através das repetições.

Interpretando-se os valores de F para blocos e \hat{c}^2 , pode-se enquadrar o presente experimento na situação (b), que é a ideal ao melhorista e, portanto, pode-se inferir que a experimentação foi adequada.

Os valores de \hat{c}^2 observados em bons experimentos em plantas perenes situam-se em torno de 0,10 (quando a herdabilidade individual estimada é da ordem de 0,30), ou seja, 10% da variação fenotípica total dentro do bloco. Assim, $\hat{c}^2 \leq 0,10$ podem ser classificados como baixos e $\hat{c}^2 > 0,10$, como altos, permitindo, assim, alguma inferência sobre a variabilidade espacial dentro dos blocos. Detalhes sobre a eficiência da análise espacial e sua relação com c^2 e com a herdabilidade são apresentados no tópico 14.19.

O parâmetro ρ_g é útil na inferência sobre a interação materiais genéticos x blocos, revelando que tanto menor é a interação quanto maior for ρ_g . Por extensão, tal parâmetro permite inferir, também, sobre a interação genótipo x ambiente dentro do próprio sítio ou fazenda de plantio. Imaginando que toda a área de plantio em uma fazenda comporta um grande número de blocos diferentes ($b \rightarrow \infty$), o interesse do melhorista é indagar sobre a capacidade de a média de um genótipo sobre os b blocos do experimento correlacionar-se com a média dos mesmos genótipos sobre os $b \rightarrow \infty$ blocos, ou seja, com o valor genotípico real do indivíduo.

Esta correlação é dada por $r_{\hat{gg}\infty} = [r_{ggd}]^{1/2} = \left[\frac{b\rho_g}{1+(b-1)\rho_g} \right]^{1/2}$, em que r_{ggd}

foi definida no tópico 19.10 como a correlação genética dos materiais genéticos dentro de um sítio ou local e equivale a $\hat{r}_{ggd} = \frac{b\hat{\rho}_g}{1+(b-1)\hat{\rho}_g} = 0,8634$, no presente

caso. Neste exemplo, a interação genótipo x blocos não foi significativa e a correlação ρ_g apresentou magnitude moderada, fato que deverá concorrer para uma pequena redução no ganho genético realizado, devido à ocorrência de interação genótipo x ambiente.

Em realidade r_{ggd} é um coeficiente de determinação da média dos materiais genéticos através dos blocos e pode ser usado no cômputo do ganho genético com a seleção em um local, por meio da expressão:

$$G_s = d_s r_{ggd} h_{mc}^2$$

$$= K r_{ggd} h_{mc}^2 \sigma_{Fm}$$

A acurácia seletiva, no caso, é dada por $r_{\hat{gg}} = [r_{ggd} h_{mc}^2]^{1/2}$.

No presente exemplo, selecionando-se 10% ($k = 1,755$) dos clones têm-se:

$$\hat{\sigma}_{Fm}^2 = \frac{Q_2}{nb} = \frac{0,71648}{12} = 0,0597$$

$$h_{mc}^2 = 1 - (1/F) = 1 - (1/1,7724) = 0,4358$$

$$\hat{G}_{snc} = 1,755 \cdot 0,4358 \cdot (0,0597)^{1/2} = 0,1869: \text{estimativa do ganho genético não corrigido para a perda devida à interação genótipo x ambiente.}$$

$\hat{G}_{sc} = 1,755 \cdot 0,8634 \cdot 0,4358 \cdot (0,0597)^{1/2} = 0,1614$: estimativa do ganho genético corrigido para a perda devida à interação genótipo x ambiente.

$\hat{r}_{gg} = [0,8634 \cdot 0,4358]^{1/2} = 0,6134$: estimativa da acurácia seletiva.

Assim, o ganho corrigido equivale a 86,34% do ganho não corrigido e reflete a perda devida à interação genótipo x ambiente dentro de locais.

Segundo Simmonds (1989), para seringueira, a produção dos clones em plantios comerciais é sempre inferior à produção nos ensaios. Este fato pode ser atribuído a perdas decorrentes da interação genótipo x ambiente. As metodologias aqui estudadas podem contribuir para melhor inferência sobre a produtividade comercial esperada dos clones. As estimativas de componentes de variância necessárias na aplicação destes métodos não precisam ser estimadas via análise de variância; podem, também, ser obtidas via metodologia de modelos mistos.

Os procedimentos apresentados neste tópico são igualmente aplicáveis aos testes de progêneres e não apenas aos testes clonais.

Capítulo 20

**Limites Seletivos
e Estruturas de Populações**

20.1. Limites Seletivos

O sucesso de um programa de melhoramento genético está condicionado à utilização e ao manejo corretos dos recursos genéticos ao longo dos ciclos seletivos. Assim, partindo-se de uma população com base genética ampla, devem ser dispendidos esforços na obtenção de bons progressos genéticos imediatos, por meio da exploração da variabilidade genética livre, expressa em cada ciclo, ao mesmo tempo em que se objetiva atingir o teto seletivo da população, por meio do melhoramento no longo prazo. Entende-se por teto ou limite seletivo o platô de produtividade a ser atingido após vários ciclos seletivos por meio da exaustão de toda a variabilidade genética contida nas populações, ou seja, após a elevação das freqüências alélicas dos alelos favoráveis em direção a 1, partindo-se das freqüências originais das populações não melhoradas. O progresso genético no longo prazo e, consequentemente, a obtenção do limite seletivo, depende, sobretudo, da variabilidade genética potencial, isto é, daquela que é mantida ao final de cada ciclo seletivo e é liberada nos ciclos subsequentes por meio da recombinação genética. Depreende-se, então, que a obtenção do limite seletivo está intimamente relacionada com a manutenção de adequado tamanho efetivo populacional ao final de cada ciclo seletivo, como forma de evitar a perda de alelos favoráveis ao longo dos ciclos de melhoramento. Outro fator que interfere negativamente no melhoramento no longo prazo é a ocorrência de endogamia devida ao pequeno tamanho efetivo populacional, a qual contribui para a atenuação dos ganhos obtidos com o melhoramento, por meio do efeito da depressão endogâmica (somente para caracteres exibindo dominância alélica) reduzindo a média populacional. Assim, limites seletivos, tamanho efetivo populacional e endogamia são intimamente relacionados no contexto de um programa de melhoramento genético.

A teoria dos limites seletivos foi estabelecida inicialmente por Robertson (1960) sob a óptica da probabilidade de fixação dos alelos favoráveis ao longo do processo seletivo.

A obtenção do limite seletivo está associada ao avanço genético (Δ_p), definido como o desvio da freqüência alélica final do alelo favorável em relação à sua freqüência original p . A freqüência final depende da probabilidade de fixação do alelo favorável, denominada $\mu(p)$ a qual é dependente de uma série de fatores, sob controle ou não do melhorista, conforme será descrito a seguir.

Assim, tem-se $\Delta_p = \mu(p) - p$, em que (Pereira & Vencovsky, 1988):

$$\mu(p) = \frac{1 - \exp [-2N_e(KE/2C)p]}{1 - \exp [-2N_e(KE/2C)]}$$

N_e = tamanho efetivo populacional.

K = diferencial de seleção praticado.

$E = \frac{2\mu}{\bar{X}}(100)$ = diferença entre os genótipos homozigotos, expressa em porcentagem da média (\bar{X}) da população.

$C = \frac{\sigma}{\bar{X}}(100)$ = coeficiente de variação da população.

Porém, a quantidade μ não pode ser avaliada diretamente em programas de melhoramento. Entretanto, Pereira & Vencovsky (1988) estimaram μ a partir de estimativas da variância genética aditiva do caráter de interesse e supondo variados números de locos.

A expressão apresentada adequa-se às seguintes condições: (a) seleção massal, entretanto, pode ser usada de modo mais amplo; (b) ausência de dominância, mas, em casos de dominância, a probabilidade de fixação é maior, de forma que a expressão apresentada é conservadora (subestimada); (c) um loco com dois alelos, podendo ser extrapolada para vários locos por meio de processo dedutivo; (d) ausência de ligação e epistasia.

Assim, esta expressão pode ser tomada como básica para o estudo dos fatores que afetam o limite seletivo. Da mesma expressão, pode-se deduzir que a obtenção do teto seletivo para um caráter quantitativo está na dependência: do tamanho efetivo populacional mantido; da intensidade de seleção praticada; da herdabilidade associada ao caráter; das freqüências iniciais dos alelos favoráveis na população; do número de locos que controla o caráter. Destes fatores, pode-se verificar que dois deles (tamanho efetivo e intensidade de seleção) estão sob total controle do melhorista, um (número de locos) está fora do controle do melhorista e os outros dois (herdabilidade e freqüências iniciais) estão sob controle parcial do melhorista, já que este pode melhorar a precisão experimental na avaliação dos genótipos e iniciar seus programas de melhoramento a partir de germoplasmas com maiores médias.

Pereira & Vencovsky (1988) concluíram que, destes fatores, o que mais influencia a obtenção do limite seletivo é o tamanho efetivo populacional, explicando 46,8% da variação observada em Δ_p , seguido das freqüências iniciais dos alelos (22,4%) e do número de locos (9%). Verificou-se, assim, que o tamanho efetivo é preponderante e, consequentemente, manter um N_e adequado é suficiente para suprir a inadequação dos demais fatores. Assim, os autores propuseram a expressão N_{eh} (onde h é a raiz quadrada da herdabilidade) como indicador do N_e que deve ser mantido em função das diferentes herdabilidades observadas. Com base nesta expressão e considerando um caráter com grande número de locos (100), populações com diferentes níveis de melhoramento e alta intensidade de seleção (2%), pode-se estabelecer a Tabela 20.1 a seguir, que fornece o N_e ideal a ser mantido em populações de melhoramento.

Tabela 20.1. Tamanho efetivo (N_e) a ser mantido em função de diferentes herdabilidades (h^2) e graus de melhoramento das populações.

	h^2	N_e a ser mantido
Populações pouco melhoradas $(\bar{p} = 0,33)$ $(N_e \cdot h = 22)$	5%	100
	10%	70
	20%	50
	30%	40
	40%	35
	50%	31
Populações com nível intermedio de melhoramento $(\bar{p} = 0,50)$ $(N_e \cdot h = 17)$	5%	75
	10%	54
	20%	40
	30%	31
	40%	27
	50%	24
Populações melhoradas $(\bar{p} = 0,67)$ $(N_e \cdot h = 9)$	5%	40
	10%	29
	20%	20
	30%	16
	40%	14
	50%	13

* $N_e \cdot h$ – obtidos a partir de dados apresentados por Pereira & Vencovsky (1988).

A Tabela 20.1 fornece bastante segurança quanto ao N_e a ser adotado, pois permite adotar N_e 's diferentes em função do grau de melhoramento das populações e das variações nos valores da herdabilidade decorrentes das diferentes condições experimentais e dos diferentes caracteres. É, particularmente, importante para o melhoramento de plantas perenes, no qual, via de regra, a seleção é baseada no procedimento BLUP associado a altas intensidades de seleção. Assim, a intensidade de seleção (2%) utilizada para estabelecimento da Tabela 20.1 é bastante adequada.

É importante frisar que para populações de 1^a geração (populações pouco melhoradas) os N_e 's deverão ser praticados de acordo com o terço superior da Tabela 20.1, enquanto que os demais terços deverão passar a ser utilizados de conformidade com o avanço das gerações.

Outros estudos realizados atingiram resultados similares aos já apresentados. Rawlings (1970) mostra que um N_e ao redor de 30 seria adequado para a maioria dos sistemas genéticos. Baker & Curnow (1969), trabalhando com 150 locos independentes, herdabilidade de 20% e intensidade de seleção de 4%, concluiu que com $N_e = 64$, 92% do ganho (seleção massal) obtido mantendo-se N_e infinito, seria alcançado. Este resultado aproxima-se daquele obtido a partir de Pereira & Vencovsky (1988) (para populações pouco melhoradas), que trabalharam com 95% de probabilidade de fixação, porém com intensidade de seleção maior (2%). Kang (1979) realizou relevante estudo, similar ao desses autores, utilizando um modelo aditivo baseado em seleção massal, 100 locos inde-

pendentes, herdabilidades de 20%, 95% de probabilidade de fixação de alelos favoráveis e variáveis freqüências iniciais dos alelos favoráveis. Os resultados estão resumidos na Tabela 20.2.

Tabela 20.2. Tamanho efetivo (N_e) a ser mantido em função de diferentes proporções de seleção e freqüências alélicas iniciais das populações.

Freqüência alélica inicial	N_e a ser mantido				
	1%	5%	10%	25%	50%
0,01	281	364	426	590	937
0,05	56	73	85	118	187
0,10	28	36	43	59	94
0,25	11	15	17	24	38
0,50	6	7	8	12	18
0,75	3	4	5	6	10

Adaptada de Kang (1979).

Constata-se, assim, que os resultados obtidos a partir de Pereira & Venkovsky (1988) são bastante conservadores em relação aos relatados por Kang (1979) e, portanto, mais seguros. A abordagem daqueles autores apresenta, também, vantagem adicional, pois permite prever o valor de N_e em função da variação da herdabilidade. Assim, com base naqueles resultados, verifica-se que, de maneira geral, para caracteres com baixa herdabilidade ($\leq 10\%$), valores de N_e de 30 a 100 são adequados.

Uma questão que, geralmente, interessa o melhorista é saber o número de gerações de seleção necessário para a obtenção do limite seletivo. A esse respeito, Robertson (1960) afirma que são necessárias $1,4 N_e$ gerações para conseguir metade (meia vida) do avanço genético possível até a obtenção do limite. Para um N_e de 60, mantido a cada geração, seriam necessárias, então, 84 gerações de seleção para atingir metade do limite seletivo. Logicamente, em um programa aplicado, dificilmente se realizará seleção em tantas gerações. Isto se deve ao fato de que os ganhos por geração se tornarão extremamente baixos, como decorrência da progressiva diminuição da variabilidade genética, bem como se pode não conseguir precisão adequada na seleção e os retornos poderão não compensar o custo operacional do melhoramento.

O ganho genético total até a obtenção do limite seletivo é dado por:

$$G_S = 2 N_e K h_f^2 \sigma_F \text{ ou } G_S = 2 N_e K h_f^2 \sigma_F \left[1 - \left(1 - \frac{1}{2 N_e} \right)^g \right] \text{ para situações de } N_e \text{ baixo, um}$$

modelo com infinitos locos e ação gênica aditiva (Robertson, 1960). Assim, o ganho genético máximo é diretamente proporcional ao N_e , mas com o aumento deste, o número (g) de gerações para se conseguir tal ganho cresce muito e, ao mesmo tempo, os ganhos por geração (ganhos imediatos) tornam-se ínfimos.

Interessante, também, é notar a influência da intensidade de seleção na obtenção do limite seletivo. Conforme observado a partir de Kang (1979), proporções de seleção mais altas (intensidades de seleção mais fracas) demandam maior tamanho efetivo e vice-versa. Também, sabe-se que os dois fatores são antagônicos, ou seja, praticando-se alta intensidade de seleção, reduz-se drasticamente o N_e . Torna-se imperativo, então, associar a estes dois fatores, o coeficiente de herdabilidade. Uma vez que mais altas herdabilidades e intensidades de seleção (menor proporção de seleção) demandam menor N_e , conclui-se que altas intensidades de seleção devem ser usadas e deve-se trabalhar procurando aumentar o coeficiente de herdabilidade, através da diminuição da variância ambiental. Os benefícios do aumento da herdabilidade são evidentes, pois com maior precisão na seleção, a adoção de alta intensidade de seleção pouco deverá contribuir para a perda aleatória de alelos favoráveis.

Notadamente em espécies perenes, que via de regra apresentam ciclos reprodutivos longos, a utilização de alta intensidade de seleção visando à obtenção de altos progressos genéticos imediatos (curto prazo) é uma necessidade. Também para a obtenção da própria fixação é necessário explorar a variabilidade genética livre, manifestada em cada geração, pois sem pressão de seleção adequada, dificilmente se caminhará em direção à obtenção do limite seletivo. Robertson (1961) afirma que para conseguir progresso genético máximo, no longo prazo, uma intensidade de seleção ideal seria 50%. Entretanto, com essa intensidade de seleção e uma herdabilidade de 20%, pode ser obtido a partir de (Nicholas, 1980, eq. 14) que seriam necessárias cerca de 460 gerações para se atingir a metade do limite, o que é impraticável.

Em suma, pode-se dizer que a intensidade de seleção deve ser alta e o tamanho efetivo populacional pode ser baixo, desde que se trabalhe com boa precisão na seleção.

Outros enfoques com relação ao tamanho efetivo em populações de melhoramento podem ser especulados, como o tamanho efetivo necessário para a obtenção de razoáveis ganhos em um número (5 a 10) limitado de gerações, bem como o N_e adequado à preservação de alelos neutros à seleção visando garantir a flexibilidade das populações em face da mudança nos objetivos da seleção. No primeiro caso, obtém-se de Barker & Curnow (1969) e Mahalovich & Bridgwater (1989) que $N_e = 32$ e 48 , respectivamente, são suficientes para a obtenção de ganhos genéticos adequados durante 10 gerações. O número de Mahalovich & Bridgwater (1989) foi maior, provavelmente, devido ao fato de esses autores trabalharem com seleção de família (não com seleção massal) e cruzamentos preferenciais positivos. Logicamente que, trabalhando-se com a abordagem dos limites seletivos, estes números (32 e 48) são automaticamente atendidos.

No caso da preservação de alelos raros, para alelos com freqüência de 1%, Kang (1979) conclui que um N_e de 160 garante a presença do alelo após 30 gerações. De acordo com Resende & Vencovsky (1990) é preciso $N_e = 75$ para preservação de alelos com freqüência de 5%. Para alelos mais raros, maiores valores de N_e são necessários.

De tudo que foi apresentado, percebe-se que uma adequada comparação de métodos de melhoramento deveria levar em conta o limite seletivo. Como a probabilidade de fixação de alelos favoráveis é proporcional a $N_e s$ e s (coeficiente seletivo) é inerente ao caráter e , portanto, constante para os diferentes métodos de melhoramento, basta considerar nas comparações apenas o N_e e o progresso genético por unidade de tempo.

Finalmente, é preciso dizer que a generalização da teoria dos limites seletivos demanda cautela. Isto se deve à complexidade e ao grande número de fatores que atuam sobre a variabilidade genética de um caráter quantitativo. Dentro desses, os principais são a mutação, que de fato contribui para a manutenção de variância antes que o limite seja atingido, e a seleção natural, que pode atuar de maneira oposta e concomitantemente à seleção artificial, de forma a atenuar os progressos genéticos, mesmo que ainda exista variabilidade genética (Hill & Caballero, 1992). Assim, existe a possibilidade de obtenção de um platô de produtividade ainda com alto nível de variabilidade genética. Quanto às mutações, Hill & Rasbash (1986) afirmam que um N_e maior que 50 garante contínua resposta à seleção, nos casos em que mutações recorrentes possam gerar variabilidade. Conforme Hill & Caballero (1992), as forças mantenedoras da variabilidade das populações, precisam ser melhor compreendidas, visando entender completamente os fatores que governam a resposta à seleção artificial.

20.2. Seleção com Restrição ao Incremento da Endogamia

Tem sido demonstrado que a utilização da metodologia de modelos mistos apresenta propriedades ótimas com respeito apenas ao ganho genético após uma geração de seleção (Henderson, 1975). Os resultados de simulação (Dempfle, 1975 e 1990; Belonksy & Kennedy, 1988; Toro & Perez-Enciso, 1990; Verrier et al., 1993) e de predição determinística (Verrier et al., 1989) indicam que o uso de pesos ótimos às informações de parentes (indivíduos e famílias de acordo com a teoria dos índices de seleção e BLUP) não conduzem a um ganho genético máximo no longo prazo. Existe, assim, um conflito entre ganho genético no curto e longo prazos, em termos de método de seleção. Tem sido demonstrado, também, que os índices de seleção conduzem a grande redução da diversidade genética (Quinton et al., 1992), principalmente se a seleção for restrita quanto ao N_e (Resende & Bertolucci, 1995). A ênfase dada à informação de família resulta em alta correlação intraclasses entre valores genéticos preditos dos indivíduos selecionados, reduzindo-se o N_e .

Para contornar esta questão, o melhorista pode atuar basicamente em três frentes: (i) a adoção de métodos de seleção que conduzam a um maior ganho genético acumulado no longo prazo; (ii) a adoção da seleção com base em índice de seleção ou BLUP, porém impondo-se restrições ao tamanho efetivo de forma a restringir o incremento da endogamia; (iii) a adoção de estruturas de populações que permitam conciliar os objetivos de máximo ganho genético no curto, médio e longo prazos.

Quanto aos métodos de seleção, o método de seleção dentro de progênie propicia maior ganho genético no longo prazo (Robertson, 1960; De-

mpfle, 1975 e 1990; Young & Skavaril, 1976) do que a seleção baseada em índice de seleção ou BLUP. Robertson (1960) concluiu que a seleção dentro de progênies e a seleção massal permitem a obtenção de iguais limites seletivos, pois, com a seleção dentro de famílias, o tamanho efetivo é praticamente duplicado, mas não se utiliza toda a variação genética aditiva. Entretanto, Dempfle (1975) mostrou que a seleção dentro de família implica menor taxa de decréscimo da variação genética do que a seleção massal, resultando em maior limite seletivo. Resultados experimentais em populações de ratos mostraram que a seleção massal conduziu a um platô seletivo após 18 gerações de seleção, ao passo que a seleção dentro de família continuou responsiva (Lopez-Fanjul, 1989). Entretanto, restringir a seleção para explorar apenas a variação dentro de progênies reduz seriamente a precisão na seleção. Dessa forma, parece adequado maximizar a precisão via utilização do BLUP, porém impondo restrições ao tamanho efetivo, não se adotando a seleção truncada.

A imposição de restrições à seleção baseada nos valores genéticos preditos pelo procedimento BLUP, a fim de manter um tamanho efetivo mínimo necessário, pode ocorrer por meio das seguintes estratégias: (*i*) redução da intensidade de seleção; (*ii*) restrição no número de indivíduos selecionados por família; (*iii*) restrição na variância do número de indivíduos selecionados por família. Em geral, as duas últimas estratégias levam a melhores resultados, maximizando o ganho genético para a condição de N_e restrito (Toro & Perez-Enciso, 1990; Resende & Bertolucci, 1995). A própria redução no número de indivíduos selecionados por família para valores baixos ($n \leq 5$) já conduz a uma diminuição na variância do número de indivíduos selecionados por família, aumentando o N_e , sem reduzir drasticamente o progresso genético (Resende & Bertolucci, 1995).

Os aspectos concernentes às estruturas de populações e seus efeitos no tamanho efetivo populacional, na endogamia e diversidade genética no longo prazo, são apresentados no tópico seguinte.

20.3. Estrutura de Populações de Melhoramento

Populações que não se enquadram na definição teórica de panmíticas (grandes e sob cruzamentos aleatórios) são ditas populações estruturadas. A estruturação pode se dar por meio de subpopulações com limitada migração e fluxo gênico entre elas até subpopulações totalmente isoladas.

Em melhoramento genético, estrutura de populações refere-se à subdivisão das populações de melhoramento como forma de atender aos seus objetivos de propiciar ganhos genéticos no curto (população de produção de propágulos melhorados), médio (população de melhoramento) e longo prazos (população de conservação).

Uma estrutura hierárquica unidirecional em três níveis é prática comum em programas de melhoramento genético animal em que os níveis são denominados núcleo, multiplicador e base (Robertson & Asker, 1951; Robertson, 1953; Bichard, 1971; Willis, 1991). O núcleo refere-se aos animais elite do programa

de melhoramento, o multiplicador destina-se à multiplicação da superioridade genética obtida no núcleo e a base é a população comercial. Esta estrutura evoluiu com os trabalhos de Jackson & Turner (1972) e James (1977, 1978) que demonstraram as vantagens de abrir os níveis superiores à migração de indivíduos superiores dos níveis inferiores. Essas vantagens são maior ganho genético e menor taxa de endogamia nos níveis superiores. Estes autores implementaram uma estrutura em dois níveis. Recentemente, uma estrutura aberta em três níveis foi delineada (Shepherd & Kinghorn, 1992; Shepherd, 1997) para o melhoramento animal. Para o melhoramento florestal, uma estrutura de três níveis foi proposta e estudada por Resende (1999b).

A principal estrutura de população proposta em literatura é a estrutura em dois níveis denominada “núcleos de melhoramento ou cruzamento” e foi o princípio dirigido ao melhoramento de carneiros (Jackson & Turner, 1972; James, 1977, 1978). Esta estrutura baseia-se na divisão da população de melhoramento em duas subpopulações: (i) núcleo de cruzamentos ou populações-elites; (ii) população principal. O núcleo constitui-se dos indivíduos com os maiores valores genéticos enquanto a população principal contém o restante dos indivíduos. Os dois segmentos são selecionados para o mesmo caráter objetivo e não existe restrição quanto à alocação de indivíduos parentes para qualquer um dos segmentos. Maiores esforços em cruzamento e teste são despendidos para a subpopulação do núcleo, podendo até mesmo serem adotados diferentes delinearmentos de cruzamento para o núcleo e para a subpopulação principal. Desse forma, este tipo de estrutura enfatiza o maior uso de indivíduos com os maiores valores genéticos.

20.3.1. Núcleo fechado de melhoramento

Neste tipo de estrutura, a cada geração, alguns indivíduos do núcleo são transferidos à subpopulação principal visando aumentar a sua taxa de melhoramento (Fig. 20.1).

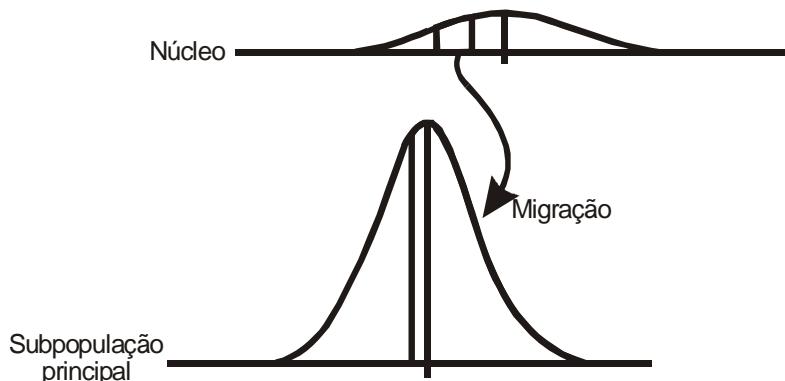


Fig. 20.1. Núcleo fechado de melhoramento.

Verifica-se que a subpopulação principal apresenta alguns indivíduos com maiores valores genéticos do que os piores indivíduos do núcleo.

21.3.2. Núcleo aberto de melhoramento

Neste tipo de estrutura, indivíduos da subpopulação principal também são transferidos ao núcleo, com o objetivo de injetar variabilidade genética e amenizar a endogamia (Fig. 20.2).

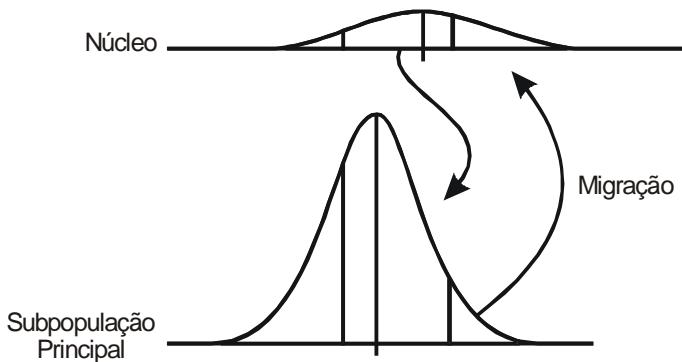


Fig. 20.2. Núcleo aberto de melhoramento.

Esse sistema conduz, ainda, a maior ganho genético em toda a população, sobretudo no núcleo. O ganho extra em relação ao núcleo fechado é da ordem de 5% a 15% variando de acordo com as intensidades de seleção (Cardellino & Rovira, 1987; Kinghorn, 1992).

Entretanto, no longo prazo, os dois tipos de estrutura (aberto e fechado) podem conduzir a um parentesco elevado entre as duas subpopulações, sendo que no esquema aberto o parentesco cresce mais rapidamente.

A estrutura de núcleos abertos tem sido utilizada com sucesso no melhoramento animal e florestal (James, 1989; Cotterill, 1989; Cotterill et al., 1989).

20.3.3. Estrutura bidirecional e hierárquica em três níveis

Estruturas de melhoramento em três níveis existem em alguns sistemas de produção de ovinos (Carrick & England, 1990) e receberam um tratamento mais formal por Shepherd & Kinghorn (1992) e Shepherd (1997). A estrutura em três níveis proporciona redução na taxa de endogamia e aumento no tamanho efetivo total quando comparada à estrutura em dois níveis. Este efeito benéfico advém da proporção assintótica de alelos contribuídos pelos diferentes níveis. A Fig. 20.3 ilustra este tipo de estrutura.

Para cálculo do N_e , na estrutura em três níveis, inicialmente, deve-se determinar a proporção assintótica de alelos contribuídos por cada nível. Definindo H como a matriz da transição alélica entre níveis, em que h_{ij} representa a proporção de alelos no nível i ao tempo t , o qual originou-se do nível j no tempo de $t-1$, tem-se (Shepherd, 1997):

$$H = \begin{bmatrix} 1 - \mu_1 - \mu_3 & \mu_1 & \mu_3 \\ d_1 & 1 - \mu_2 - d_1 & \mu_2 \\ d_3 & d_2 & 1 - d_2 - d_3 \end{bmatrix}, \text{em que:}$$

μ_i = taxa de migração entre níveis, no sentido ascendente.

d_i = taxa de migração entre níveis, no sentido descendente.

O fluxo de migração pode ser melhor visualizado por meio da Fig. 20.3:

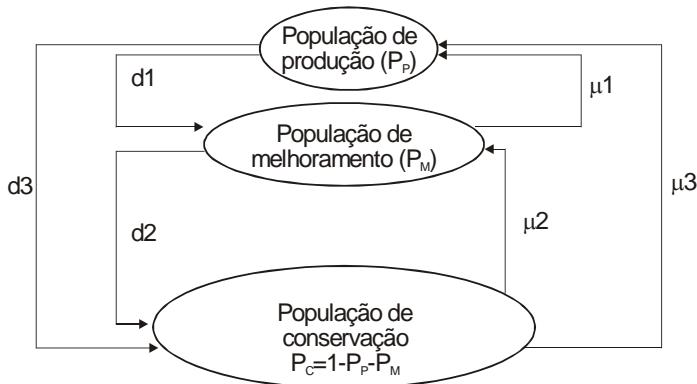


Fig. 20.3. Estrutura de populações em três níveis, em que μ_i e d_i referem-se às taxas de migração entre os níveis, nos sentidos ascendente e descendente, respectivamente e P_p , P_m e P_c são relativos às frações da população total cruzadas no núcleo (população de produção), população de melhoramento e de conservação, respectivamente.

A matriz H é análoga à matriz de fluxo gênico apresentada por Hill (1974). Pode ser demonstrado (Hill, 1974) que H tem um único autovalor unitário e os outros apresentam menores valores absolutos. Para o autovalor unitário, um autovetor à direita é $r'_1 = (1 1 1)$ e um autovetor à esquerda é $\ell'_1 = (\ell_1 \ell_2 \ell_3)$, em que:

$$\ell_j = \lambda_j / \sum_{j=1}^3 \lambda_j$$

$$\lambda_1 = \mu_2 d_3 + d_1 d_2 + d_1 d_3$$

$$\lambda_2 = \mu_1 d_2 + \mu_1 d_3 + \mu_3 d_2$$

$$\lambda_3 = \mu_1 \mu_2 + \mu_2 \mu_3 + \mu_3 \mu_1$$

À medida que t aumenta, pode ser demonstrado que:

$$H' \rightarrow r_1 \quad \ell'_1 = \begin{bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell_1 & \ell_2 & \ell_3 \\ \ell_1 & \ell_2 & \ell_3 \end{bmatrix}$$

Então, ℓ_j ($= \ell_i$) representa a proporção assintótica de alelos contribuída para cada nível. A proporção de alelos contribuída por cada nível flutuará inicialmente, mas com o tempo estabilizará para estes valores assintóticos.

O tamanho efetivo da população composta total equivalerá, então, a (Sheperd, 1997):

$$\frac{1}{N_{ec}} = \ell_1^2 \frac{1}{N_{e1}} + \ell_2^2 \frac{1}{N_{e2}} + \ell_3^2 \frac{1}{N_{e3}}, \text{ em que:}$$

N_{ei} refere-se ao tamanho efetivo de cada nível fechado e, ℓ_i são funções das taxas de migração.

A taxa assintótica (F_a) de endogamia em relação à população principal fechada equivale a:

$$F_a = \frac{N_{e2}}{N_{ec}} = \ell_1^2 \frac{N_{e2}}{N_{e1}} + \ell_2^2 \frac{N_{e2}}{N_{e3}} + \ell_3^2 \frac{N_{e2}}{N_{e3}}$$

Considerando, por exemplo, os tamanhos efetivos $N_{e1} = 10$; $N_{e2} = 50$ e $N_{e3} = 100$ para o núcleo, multiplicador (população de melhoramento) e base (população de conservação), respectivamente, e $\mu_1 = 0,30$; $\mu_2 = 0,20$; $\mu_3 = 0,20$; $d_1 = 0,30$; $d_2 = 0,10$; $d_3 = 0,01$, tem-se $N_{ec} = 102$, que equivale a 10 vezes o N_e do núcleo e 2 vezes o N_e do multiplicador. Verifica-se, assim, a vantagem deste tipo de estrutura em relação às estruturas tradicionais.

20.3.4. Comparação das três estruturas no melhoramento de espécies perenes

Uma consideração importante na comparação entre as três estruturas é a possibilidade de ocorrência de interação genótipo x ambiente. Considerando a existência de dois ambientes, a utilização de dois núcleos fechados é melhor se a correlação genética entre o desempenho nos dois ambientes for inferior a 0,80, enquanto que, se cada população possui seu próprio ambiente, um sistema de núcleos abertos é melhor desde que a correlação genética média entre desempenho nos vários ambientes for superior a 0,55 (Del-Bosque-Gonzales & Kinghorn, 1987). Uma estrutura em três níveis, otimizada, permite uma taxa de ganho genético mais rápido e fornece o limite superior do progresso genético possível com a estrutura de núcleos abertos (James, 1989).

Uma revisão detalhada sobre tamanho e estruturas de populações de melhoramento foi apresentada por Resende (1999a). Dentre as principais estru-

ras de população empregadas no melhoramento florestal destacam-se: núcleo de cruzamento (Cotterill, 1989; Cotterill et al., 1989), sublinhas (Van Buijtenen & Low, 1979) e multiplicações (Namkoong et al., 1980). Estas três estruturas constituem opções razoáveis para otimização do ganho genético no curto e médio prazos por permitirem enfatizar o uso dos indivíduos com os maiores valores genéticos e gerenciar a endogamia na população de produção comercial.

A estrutura hierárquica em três níveis, proposta por Resende (1999b), considera com maior rigor a questão dos limites seletivos e suas consequências para o melhoramento no longo prazo. Tal estrutura considera os tamanhos efetivos $N_{e1} = 10$, $N_{e2} = 50$ e $N_{e3} = 100$ para as populações do núcleo, principal (melhoramento) e de conservação (baseada na seleção dentro de progénies, de apenas um indivíduo em cada). Adicionalmente considerou-se que o núcleo a cada geração é formado por 50%, 30% e 20% de indivíduos do próprio núcleo, da população principal e da população de conservação, respectivamente, e que a população principal a cada geração é formada por 50%, 30% e 20% de indivíduos da própria população principal, do núcleo e da população de conservação, respectivamente. Neste caso, $\mu_1 = 0,30$, $\mu_2 = 0,20$, $\mu_3 = 0,20$, $d_1 = 0,30$, $d_2 = 0$ e $d_3 = 0$. Valores alternativos de d_2 e d_3 foram também avaliados com a finalidade de verificar os efeitos da manutenção da população de conservação isolada. Assim, foram obtidos os valores de N_{ec} e F_a , apresentados na Tabela 20.3

A estrutura em três níveis apresentada por Shepherd (1997) difere da estrutura proposta por Resende (1999b) por permitir a transferência de indivíduos dos demais níveis para a base. Logicamente, as duas estruturas em consideração diferem também em termos de métodos ou unidades de seleção utilizados.

Para a estrutura piramidal em três níveis, são apresentados, na Tabela 20.3, os valores do tamanho efetivo da estrutura total, bem como os valores da taxa assintótica de endogamia em relação a população principal fechada.

Tabela 20.3. Tamanho efetivo (N_{ec}) da estrutura de população em três níveis bem como taxa assintótica de endogamia (F_a) em relação à população principal fechada ($N_e = 50$) e funções das taxas de migração (ℓ_i) para variadas taxas de migração (d) da população principal (d_2) e do núcleo (d_3) para a população de conservação.

d_2	d_3	N_{ec}	F_a	ℓ_1	ℓ_2	ℓ_3
0,30	0,10	80	0,63	0,29	0,38	0,33
0,10	0,05	122	0,41	0,20	0,23	0,57
0,10	0,02	99	0,51	0,24	0,38	0,38
0,10	0,01	102	0,49	0,23	0,38	0,39
0,09	0,000	146	0,34	0,12	0,19	0,69
0,05	0,02	138	0,36	0,12	0,14	0,74

Continua...

Tabela 20.3. Continuação.

d_2	d_3	N_{ec}	F_a	ℓ_1	ℓ_2	ℓ_3
0,05	0,01	139	0,36	0,10	0,13	0,77
0,05	0,001	138	0,36	0,07	0,13	0,80
0,02	0,02	129	0,39	0,08	0,08	0,84
0,02	0,01	125	0,40	0,06	0,07	0,87
0,02	0,001	119	0,42	0,04	0,06	0,90
0,01	0,01	117	0,43	0,05	0,05	0,90
0,01	0,001	110	0,45	0,02	0,03	0,95
0,001	0,001	102	0,49	0,005	0,005	0,99
0,000	0,000	100	0,50	0,000	0,000	1,00
		160	0,31	0,0625	0,3125	0,625

Constata-se pela Tabela 20.3 que a manutenção da população de conservação totalmente isolada ($d_2 = d_3 = 0$) mantém o N_e total da estrutura em 100, fato que é bastante razoável pois conduz a uma taxa (F_a) assintótica de endogamia de 0,50 em relação à estrutura tradicional (população principal fechada e $N_e = 50$). Os valores de F_a denotam a magnitude da endogamia que será gerada na estrutura total em relação à magnitude da endogamia que será gerada na população principal. Assim, com a inclusão da população de conservação na estrutura, a endogamia no longo prazo será reduzida em 50%.

Para as taxas de migração consideradas ($\mu_1 = 0,30$, $\mu_2 = 0,20$, $\mu_3 = 0,20$, $d_1 = 0,30$), pode-se aumentar o N_e total da estrutura global, atribuindo-se valores apropriados para d_2 e d_3 . Para estas condições, verifica-se pela Tabela 21.3 que o N_e total será praticamente maximizado para valores de d_2 e d_3 , próximos a 9% e 0%, respectivamente. Neste caso o N_e resultante será de 146. Na realidade, o N_e máximo possível para a presente situação é de 160, valor este que equivale ao somatório dos N_e 's de cada nível da estrutura. Para atingir o valor máximo é necessário que os ℓ 's sejam proporcionais aos respectivos N_e dos vários níveis, pois assim sendo, Resende & Vencovsky (1990) demonstraram que $N_{ec} = \sum_i N_{ei}$. Para obter $N_{ec} = 160$, é preciso ter $\ell_1 = 0,0625$; $\ell_2 = 0,3125$; e $\ell_3 = 0,625$ (Tabela 20.3).

Para as taxas de migração consideradas, tendo-se $d_2 = 0,09$ e $d_3 = 0,0$, atinge-se 91% (146/160) do N_{ec} máximo possível e 1,09% (0,34/0,31) da endogamia mínima possível, ou seja, ter-se-ia uma taxa assintótica relativa de endogamia de 0,34, ou seja, 66% a menos da endogamia em relação à endogamia da população principal tradicional ($N_e = 50$). Entretanto, neste caso, a população de conservação não seria mantida totalmente isolada e a manutenção de 100 famílias não aparentadas na população de conservação poderia ser um complicador. De qualquer forma, está claro que a transferência de indivíduos da população principal para a população de conservação só será benéfica se praticada em taxas muito baixas (Tabela 20.3).

20.3.5. Medidas de diversidade genética

Como medidas de diversidade genética associadas aos diferentes métodos de seleção e estrutura de populações podem ser considerados:

$$N_{ef} = \frac{(\sum n_i)^2}{\sum n_i^2} = \text{número efetivo de famílias selecionadas} \quad (\text{Robertson, 1961}),$$

onde n_i é o número de indivíduos selecionados na família i.

$$D = \frac{N_{ef}}{F_o} = \text{diversidade genética ou número efetivo de famílias selecionadas em relação ao número } (F_o) \text{ original de famílias} \quad (\text{Wei & Lindgren, 1996; Wei, 1996}).$$

$D = 1$ antes da seleção e $0 < D \leq 1$ após a seleção.

$$N_e = \frac{4Fn}{3+n+\sigma_n^2/n} = \text{número efetivo de indivíduos selecionados.}$$

onde F, n e σ_n^2 , referem-se ao número de famílias selecionadas, número médio de indivíduos selecionados por família e variância do número de indivíduos selecionados por família, respectivamente.

A vantagem da estrutura piramidal e hierárquica de populações, a qual inclui a população de conservação, advém do maior tamanho efetivo propiciado pela seleção dentro de progênie, o qual acarreta efeito positivo no ganho genético acumulado. Tomando a expressão geral do N_e , $N_e = 4Fn/(3+n+\sigma_n^2/n)$ verifica-se:

- Para a seleção dentro de progênies com um indivíduo selecionado por progênie, tem-se $\sigma_n^2 = 0$ e $n = 1$, de forma que $N_e = F_0$, onde F_0 é o número original de famílias. Na seleção dentro de progênies com $n = 2$, tem-se para um número total fixo de famílias a serem avaliadas na próxima geração $N_e = (F_0/2) 8/5 = (4/5) F_0$.
- Para as seleções massal e baseada no BLUP, têm-se $\sigma_n^2 > 0$ e $F < F_0$, de forma que o N_e é sempre inferior ao N_e da seleção dentro de progênies com $n = 1$.

Assim, para o melhoramento no longo prazo, as vantagens da seleção dentro de progênies sobre a seleção pelo BLUP são: maior N_e , maior ganho genético acumulado, menor endogamia por geração, ausência de seleção entre famílias.

Conclui-se, então, que para F_0 fixo e constante ao longo das gerações, a seleção de 1 indivíduo dentro de progênies propicia o maior tamanho efetivo possível. Considerando o conceito do número efetivo de famílias (Robertson, 1961) dado por $N_{ef} = 1/(\sum w_i^2)$, onde $w_i = n_i / \sum n_i$, verifica-se que este método de seleção também maximiza N_{ef} .

Na Tabela 20.4. são apresentadas estimativas (Resende 1999b) dos parâmetros descritivos da diversidade genética associada aos diferentes métodos de seleção e estrutura de populações, a partir da avaliação de 103 progênies de meios-irmãos.

Tabela 20.4. Diversidade genética (D), número efetivo (N_e) e físico (N) de indivíduos selecionados, número efetivo de famílias (N_{ef}) selecionadas e número físico de famílias (N_{ff}) selecionadas associadas a diferentes estruturas e populações-objetivo.

Estrutura	População	Método de seleção	N	N_e	N_{ff}	N_{ef}	D
Núcleo de cruzamento	Principal	BLUP	70	50	43	31,45	0,31
Hierárquica em três níveis	Conservação	Dentro de progênies	103	103	103	103	1,0

Comparando-se as estruturas de núcleo de cruzamento e hierárquica, em termos de diversidade, verifica-se que a primeira mantém apenas 31% da diversidade original. A estrutura piramidal hierárquica permite, por meio da população de conservação via seleção dentro de progênies, a manutenção de 100% da diversidade genética original. Na estrutura hierárquica, o número físico de indivíduos selecionados, o número efetivo de indivíduos selecionados, o número físico de famílias selecionadas e número efetivo de famílias selecionadas são coincidentes (Tabela 20.4).

Uma vez que, atualmente, tem-se utilizado de forma generalizada o procedimento BLUP para seleção (o qual enfatiza a informação de família), a condução, em paralelo, de uma população de conservação via seleção dentro de progênies, associada a qualquer estrutura, previne os efeitos adversos do BLUP em termos de diversidade genética, mantendo-se a viabilidade do sistema (estratégia de melhoramento no longo prazo). Outra opção é a restrição do número de indivíduos selecionados por família, conforme apresentado no tópico 20.2.

Capítulo 21

**Fundamentos de Genética
de Populações para a Conservação
de Germoplasma**

21.1. Introdução

Em muitas espécies perenes, grande parte do germoplasma jamais foi avaliado ou utilizado, tendo em vista que muitos programas de melhoramento genético iniciaram-se recentemente e que grande parte da exploração econômica concentra-se no extrativismo (casos da erva-mate, castanha-do-pará, palmito, etc.) ou são baseados em amostras restritas dos germoplasmas (caso da seringueira por exemplo). Este fato, aliado à grande devastação que vêm sofrendo as florestas naturais que constituem a área de ocorrência natural de muitas espécies de importância econômica, faz com que programas de conservação de recursos genéticos de muitas espécies perenes sejam prioritários como forma de garantir a sua utilização futura.

Do ponto de vista técnico, ações de conservação genética devem ser fundamentadas em sólidos princípios de genética de populações. Neste contexto são relevantes os conceitos de tamanho efetivo populacional e deriva genética, endogamia e depressão endogâmica, mutação e carga genética, polimorfismo e variabilidade genética, migração e fluxo gênico, subdivisão e estrutura de populações, e também de seleção.

O presente capítulo tem como objetivo elucidar os fundamentos de genética de populações necessários ao delineamento de eficientes metodologias e estratégias de amostragem e conservação de germoplasma de espécies perenes, tendo como base o trabalho de Resende et al. (1997b).

21.2. Genética de Populações e Conservação Biológica

A diversidade biológica do planeta vem sendo reduzida drasticamente como consequência direta ou indireta das ações humanas, de forma que várias espécies tornaram-se extintas e muitas outras tiveram o seu tamanho populacional reduzido, encontrando-se em perigo de extinção. Atualmente muitos esforços vêm sendo realizados com vistas à manutenção da biodiversidade, com as seguintes justificativas: valor econômico dos recursos biológicos, benefícios ecológicos e ambientais, estética e direito de existência dos seres vivos. Neste sentido, a conservação da biodiversidade é necessária em três níveis: diversidade genética dentro de espécies, diversidade de espécies e diversidade do ecossistema (Frankham, 1995a). A genética de populações está diretamente relacionada com os dois primeiros níveis de biodiversidade e, indiretamente, ao terceiro.

Os principais fatores que contribuem para a mencionada extinção são a perda de habitat, a exploração desregrada, a introdução de espécies e a poluição. Estes fatores, geralmente, reduzem o tamanho populacional das espécies, fato que as torna susceptíveis a efeitos estocásticos (com variações aleatórias) tais como variações ambientais e demográficas, catástrofes e efeitos genéticos (depressão endogâmica, perda de variação genética e acúmulo de mutações deletérias) (Frankham, 1995a). Está explícito, então, que o fator decisivo para a

extinção das espécies é o pequeno tamanho populacional, que é, em essência, um problema de genética de populações.

A conservação biológica é uma ciência multidisciplinar baseada principalmente na ecologia e na genética. Segundo Frankham (1995a) existem sete principais aspectos genéticos envolvidos na conservação biológica: depressão endogâmica; acúmulo e perda de mutações deletérias; perda de variação genética em pequenas populações; adaptação genética *ex situ* e efeitos no sucesso de reintrodução; depressão por alogamia; subdivisão de populações e redução na migração; introgessão.

Este capítulo tratará, em tópicos específicos, de alguns (exceto adaptação genética *ex situ*, depressão por alogamia e introgessão) destes aspectos. A depressão por alogamia refere-se à redução no valor adaptativo dos indivíduos como consequência da hibridação entre populações muito distintas dentro de espécies (ou entre espécies que se cruzam) e essa redução é atribuída à quebra de complexos gênicos co-adaptados. A adaptação genética *ex situ* diz respeito ao melhoramento genético para algum caráter de forma que a população não se torne mais apta a sobreviver por si só (sem a ação do homem) no ambiente natural (relevante para populações animais e vegetais, como o milho). A introgessão está relacionada com o cruzamento entre espécies afins que possa causar distúrbios genéticos nas populações naturais. A ocorrência destes três fatores é mais restrita e eles não serão considerados neste documento.

21.3. Tamanho Efetivo Populacional e Deriva Genética

Os conceitos básicos do tamanho efetivo populacional (N_e) foram apresentados no Capítulo 1. Em termos genéricos, pode-se dizer que o tamanho efetivo populacional é inerente ao tamanho genético de uma população reprodutiva e não ao número de indivíduos que a compõe. Assim, diz respeito à representatividade genética de amostras de animais, plantas e sementes. Conforme implícito no próprio conceito de Wright, na situação de um tamanho efetivo pequeno, um reduzido número de indivíduos participa efetivamente do intercruzamento com vistas à regeneração da nova população, levando à ocorrência de dois eventos: mudança aleatória das freqüências alélicas (oscilação ou deriva genética) e aumento da endogamia na nova população.

21.3.1. N_e e probabilidade de extinção

A estocasticidade genética, isto é, a endogamia e a deriva genética em pequenas populações, resulta em perda de variabilidade genética, fato que diminui a viabilidade da população. Tanto a endogamia quanto a deriva reduzem a heterozigose na população. A endogamia tende (exceto se alelos deletérios tenham sido eliminados da população por endogamia prévia) a reduzir a adaptação média da população, via depressão endogâmica e, consequentemente, a aumentar a probabilidade de extinção. Já a deriva causa a perda ou fixação de

alelos em pequenas populações. A perda aumenta, necessariamente, a homozigose, mas a adaptação pode aumentar ou diminuir, dependendo da ocorrência ou não de perda de alelos deletérios ou vantajosos. A fixação de alelos deletérios, por sua vez, quando medida sobre uma série de pequenas populações derivadas da mesma população genitora, provoca redução na adaptação média da população. Assim, tanto a endogamia verdadeira (medida dentro das populações) quanto a deriva genética (entre populações) aumentam, de fato, a probabilidade de extinção (Newman & Pilson, 1997; Ellstrand & Elam, 1993; Frankham, 1995b). Extinções têm sido observadas com taxas de endogamia extremamente baixas devido ao tamanho populacional finito, tais como a extinção de populações de *Drosophila melanogaster* com $N_e = 50$ em 210 gerações (Latter et al., 1995).

Um tratamento matemático relacionando N_e e tempo médio para extinção em populações dióicas com sobreposição de gerações é apresentado por Pollak (1980).

21.3.2. Estratégias para aumento do N_e e manutenção da variação genética

Em conservação genética é desejável a minimização da endogamia e da perda de variação genética no processo de gerenciamento genético das populações a serem conservadas. Esse processo baseia-se na retenção da heterozigose (que representa potencial evolutivo) e da diversidade alélica, que são frequentemente relacionados, embora a heterozigose seja menos afetada por grave afunilamento devido ao pequeno tamanho populacional (Frankham, 1995a).

A proporção esperada de variação genética neutral (heterozigose H_t) retida em uma dada população após t gerações é dada por:

$$H_t / H_o = [1 - 1/(2N_e)]^t = 1 - F, \text{ em que:}$$

H_o : heterozigose inicial.

F : coeficiente de endogamia.

Por esta expressão, verifica-se que a retenção de heterozigose é maximizada por:

- a) Maximização da heterozigose inicial, trabalhando com populações com maiores variabilidades genéticas e com maiores números de fundadores não aparentados. A imigração também é uma forma efetiva de ampliação de variabilidade e aumento da adaptação reprodutiva.
- b) Maximização do intervalo entre gerações, ou seja, minimização do número de gerações. Conforme visto no Capítulo 1, há perda de representatividade genética em relação à população inicial a cada passagem de geração, mesmo que o N_e seja mantido constante em cada uma. Assim, a conservação via armazenamento de sementes por longos períodos também é desejável.

c) Maximização do tamanho efetivo populacional.

Conforme visto no Capítulo 1, o N_e depende do número (N) total de indivíduos da população reprodutora, da variação no tamanho de família, da desigual proporção de sexo, da flutuação de N através das gerações e da seleção dentro da progênie (que afeta a média e a variância (σ_k^2) do número de gametas contribuídos).

Assim, as seguintes ações contribuem para maximizar o N_e .

- a) Maximizar N e, simultaneamente, a relação N_e/N . N_e/N não deve ser maximizado apenas às custas de um baixo N .
- b) Manter igual tamanho de família ($\sigma_k^2 = 0$). Este procedimento, praticamente, duplica o N_e .
- c) Manter a proporção de sexo próxima de 1:1. Quanto maior a discrepância entre número de machos e de fêmeas, menor o N_e .
- d) Manter um N_e desejável e constante nas gerações, já que o N_e (tamanho efetivo em relação à população original) em uma determinada geração corresponde à média harmônica dos N_e mantidos nas gerações anteriores, a qual depende sobretudo dos valores baixos de N_e .

Deve-se evitar a seleção entre famílias, a qual reduz o N_e . A seleção dentro de famílias deve ser praticada em intensidades iguais dentro de famílias ($\sigma_k^2 = 0$).

Na experimentação devem ser utilizadas parcelas com poucas plantas e muitas repetições.

As estratégias apresentadas são mais adequadas à conservação genética ex situ.

21.3.3. Proporção N_e/N em populações silvestres

Resultados experimentais revelaram que as estimativas de N_e/N na natureza equivalem, em média, a 0,11, valor muito mais baixo do que geralmente suposto (Frankham, 1995c). Esta informação reforça a necessidade de cuidados amostrais por ocasião da coleta de germoplasma em populações naturais, devendo-se amostrar indivíduos espacialmente distantes. Procedendo desta forma, torna-se reduzida a possibilidade de que sejam tomados indivíduos aparentados.

O gerenciamento genético destas populações naturais pode potencialmente aumentar o N_e , por meio de intervenções humanas com o objetivo de manter a proporção de sexo em 1:1 (espécies dióicas) e N constante nas diferentes gerações. A manutenção do mesmo tamanho de família seria interessante não apenas por praticamente duplicar o N_e , mas também por remover a seleção natural entre famílias. Mas na natureza esta última prática é bastante dificultada.

Uma vez que N_e/N na natureza é baixo, o tamanho da população (N) para

conservação/preservação *in situ* deve ser muito maior do que geralmente descrito. Outro fator muito importante a ser considerado na conservação *in situ* é o fluxo gênico via bancos de sementes no solo (Martins, 1987). Maiores detalhes referentes à conservação *in situ* podem ser encontrados em Kageyama (1987, 1990).

21.3.4. Valores de N_e adequados à conservação genética

Os valores adequados de N_e a serem adotados para conservação genética são determinados em função, principalmente, de dois critérios: prevenção de depressão endogâmica e manutenção do potencial evolutivo.

Para prevenção de depressão endogâmica um N_e de 50 é suficiente. Por outro lado, para manutenção do potencial evolutivo da população indefinidamente, um N_e da ordem de 500 é suficiente, considerando o balanço entre deriva e mutação (Frankham, 1995a). Entretanto, Lande (1995) sugere que um N_e de 5.000 é necessário para reter o potencial evolutivo para o caráter adaptação reprodutiva após ajustar a taxa total de mutação omitindo as mutações severamente deletérias. Estes resultados estão resumidos na Tabela 21.1, com base em Alvarez-Buylla et al. (1996).

Tabela 21.1. Tamanhos efetivos (N_e) adequados à conservação genética considerando vários tipos de efeitos genéticos: segregação de alelos deletérios recessivos (parcialmente), perda de variação adaptativa e fixação de mutações deletérias por deriva genética.

N_e	Efeito genético	Força evolutiva	Tipo de variação genética
50*	Segregação	Endogamia	Alelos recessivos deletérios
500**	Perda de variação	Mutação, deriva	Quantitativa, todas as mutações
5000**	Perda de variação	Mutação, deriva	Quantitativa, mutações quase-neutrais (somente 10% de todas as mutações)

* Considerando 1% de aumento no coeficiente de endogamia por geração.

** Considerando o N_e na geração em que a variação genética gerada por mutação é igual à variação genética perdida por deriva genética.

Na prática, a conservação biológica enfatiza três formas de atuação (Loeschcke et al., 1994):

- Proteção de habitats em reservas, na qual a seleção de sítios e delineamentos requerem poucas informações genéticas.
- Áreas de reservas ótimas baseadas em análise de viabilidade populacional requerendo muitas informações genéticas e demográficas (Lande, 1988).
- Conservação genética *ex situ* de espécies.

Assim, a preservação *in situ* se encaixa no item b e, portanto, seria neces-

sário manter um N_e de 500-5.000 nesta situação. A preservação ex situ (item c) demandaria um N_e da ordem de 50. Entretanto, com N_e da ordem de 50, somente se conseguem reter alelos com freqüência $\geq 6\%$. Para a retenção de alelos com freqüência $\geq 1\%$ um N_e da ordem de 150 deve ser utilizado (Vencovsky, 1987b). Este número pode ser atingido com uma amostra de cerca de 40 matriizes da natureza, em populações alógamas, por meio de pelo menos 100 sementes.

21.4. Endogamia e Depressão Endogâmica

21.4.1. Conceitos de endogamia e relações com a estrutura genética de populações e sistemas reprodutivos

Conforme relatado por Malecot, os genes em uma população podem ser semelhantes por duas razões mutuamente exclusivas: (i) replicação do DNA de um único alelo existente previamente em um ancestral, portanto um é cópia do outro e surgiu por processo reprodutivo; (ii) meramente por serem semelhantes pelo fato de ambos serem a. Os primeiros são denominados idênticos por descendência e os indivíduos que os possuem são autozigóticos, e os últimos são chamados idênticos em estado e os indivíduos que os possuem são denominados alozigóticos.

O coeficiente de endogamia foi definido, inicialmente por Wright, como correlação entre gametas que se unem. Malecot o definiu como a probabilidade de os dois alelos de um loco em um indivíduo serem idênticos por descendência. Logo, está implícito que o termo endogamia refere-se ao processo de cruzamento entre indivíduos mais parentados entre si do que aqueles que são membros aleatórios da população (Kempthorne, 1957) e que tem por consequência um aumento da homozigose da população. Dessa forma, é um conceito relativo à população ancestral que também possui indivíduos homozigóticos, porém alozigóticos e não autozigóticos.

A endogamia em populações naturais pode ocorrer, principalmente, por meio de: cruzamentos entre indivíduos parentados em espécies dióicas; cruzamentos entre indivíduos parentados e autofecundações em espécies monóicas alógamas ou com sistema reprodutivo misto; autofecundações em espécies monóicas autógamas; pequeno tamanho populacional em espécies alógamas e com sistema reprodutivo misto, fato que aumenta a probabilidade de cruzamentos entre parentes e/ou autofecundação. A estrutura genética destas populações, em termos do coeficiente de endogamia e freqüências genotípicas, equivale a $(p^2+Fpq) AA + 2pq(1-F) Aa + (q^2+Fpq) aa$, em que AA, aa e Aa, são os genótipos homozigoto favorável, homozigoto desfavorável e heterozigoto, respectivamente; F é o coeficiente de endogamia e p e q são as freqüências dos alelos favorável (A) e desfavorável (a), respectivamente.

Pela expressão apresentada, para populações alógamas ($F=0$), tem-se $p^2 AA + 2pq Aa + q^2 aa$, ou seja, as proporções esperadas para populações em equilíbrio de Hardy-Weinberg. É importante frisar que, neste caso, sendo ($F=0$), os indivíduos homozigotos são alozigóticos. Para espécies autógamas tem-se ($F=1$) e, portanto, $p AA + q aa$, enquanto que para espécies com sistema reprodutivo misto, a estrutura é dependente da taxa de autofecundação (S) praticada por cada população, pois $F = S/(2-S)$, conforme definido por Li (1955).

Comparando com a população base não endógama, verificam-se, como decorrência da endogamia, o acréscimo de uma fração Fpq à freqüência de cada classe de homozigotos e um decréscimo de $2Fpq$ na freqüência dos heterozigotos.

A derivação da expressão apresentada é bastante simples. Uma população qualquer não endógama possui estrutura genotípica $(\sum p_i A_i)^2$, em que p_i é a probabilidade de um alelo qualquer desta população ser A_i . Sendo F a probabilidade de os alelos em um indivíduo serem idênticos por descendência, a probabilidade de que ambos sejam A_i é Fp_i . A probabilidade de dois alelos serem não idênticos é $(1-F)$ e a probabilidade de que dois alelos tomados ao acaso sejam A_i e A_j é $p_i p_j$. Assim, a população com estrutura $(\sum p_i A_i)^2$ submetida à endogamia medida pelo coeficiente F , apresenta estrutura genotípica:

$$F(\sum p_i A_i A_i) + (1-F)(\sum p_i A_i)^2 \text{ (Kempthorne, 1957).}$$

Considerando os alelos A e a , a população genotípica resultante será:

$$[Fp + (1-F)p^2] AA + [2(1-F)pq] Aa + [Fq + (1-F)q^2] aa ,$$

a qual caracteriza o que é denominado “equilíbrio de Wright”, ou seja equilíbrio para um valor estável de F . Observa-se por esta expressão que do homozigoto AA , a fração $(1-F)p^2$ corresponde à alozigotia e a fração Fp , à autozigotia ou endogamia propriamente dita. Está explícito, então, que o conceito de F , no que diz respeito à homozigose, é relativo a uma população base não endógama, mas que possui indivíduos homozigotos para determinados locos.

Assim como o tamanho efetivo, a endogamia apresenta vários conceitos e significados (Jacquard, 1975; Templeton & Read, 1994). Templeton & Read (1994) destacam os significados: 1) endogamia como medida de ancestralidade comum; 2) endogamia como medida de deriva genética; 3) endogamia como medida do sistema de cruzamento em uma população reprodutora.

Os dois primeiros significados estão associados aos conceitos de tamanho efetivo de endogamia e de variância, respectivamente. O terceiro significado refere-se ao coeficiente denominado índice panmítico (f), que é uma medida de endogamia como desvio de uma população de referência que possui sistema de cruzamentos totalmente aleatórios. O índice panmítico mede o sistema de cruzamento da população como um desvio da freqüência de heterozigose esperada (H_e) sob cruzamentos aleatórios, por meio da expressão $f = 1 - H_o/H_e = 1 - F$,

em que H_o é a heterozigose observada. Assim, se $f = 0$ a população está sob cruzamentos aleatórios, se $f > 0$ a população está experimentando um sistema endogámo de cruzamentos e se $f < 0$ a população está sob um sistema de cruzamento com prevenção de endogamia.

21.4.2. Endogamia devida ao pequeno N_e

Considere-se uma população de referência sendo diplóide, monóica (com autofecundação permitida), não endogáma e com tamanho N . Todos os genes em um loco são não idênticos por descendência, de forma que numa população de N indivíduos, cada loco produz, aleatoriamente, $(2N)$ diferentes tipos de gametas igualmente, com freqüência $1/(2N)$, para produzir os indivíduos da geração seguinte.

Percebe-se, então, que a probabilidade de um indivíduo ser fecundado, a partir de um gameta oriundo dele mesmo é $1/(2N)$. Assim, quanto menor for tamanho N da população reprodutiva, maior será a endogamia na geração subsequente.

O coeficiente de endogamia devido ao pequeno tamanho populacional, quantificado em termos de aumento da homozigose ou perda de heterozigose equivale a $F = 1/(2N)$, em que N diz respeito aos indivíduos não parentados na população ideal de referência, ou seja, N é igual ao número efetivo de indivíduos que participam da reprodução. Assim, de maneira generalizada, é melhor definir $F = 1/(2N_e)$, em que N_e é o tamanho efetivo populacional.

21.4.3. Efeito da endogamia na média populacional – depressão endogâmica

A média genotípica populacional com $F = 0$ equivale a $M_o = (p-q) a + 2 p q d$ e a média com endogamia F equivale $M_F = M_o - 2 p q d F$ (Falconer, 1989).

Assim, com endogamia F , a média altera-se na fração $-2 p q d F$, ou seja, é sempre reduzida, desde que haja variabilidade (p ou q não sejam iguais a zero) e que haja alguma dominância alélica (d). O fenômeno associado à redução na média é denominado depressão endogâmica, o qual resulta da mudança das freqüências genotípicas da população, após esta sofrer endogamia.

O termo depressão endogâmica caracteriza-se pelos efeitos da perda de heterozigose, comparando-se a população não endogáma (com $F = 0$) com a população totalmente endogâmica (com $F = 1$). Assim, com uma geração de autofecundação em uma população não endogáma, tem-se $F = 1/2$ e, consequentemente, metade da depressão endogâmica (D).

A depressão endogâmica pode ser utilizada para cálculo da taxa (S) de autofecundação em espécies monóicas. Considerando que a diferença entre a média de progênies de polinização cruzada (PC) e a média de progênies de polinização aberta (PA) estima $(S/2)D$ e que a diferença entre PC e a média de progênies de autofecundação (AF) estima $(1/2)D$, tem-se que $S = (PC-PA)/(PC-AF)$.

21.4.4. Efeito da endogamia na variância populacional - perda de variação genética

Considerando toda uma população, a endogamia aumenta a variância genética total, embora diminua a diversidade alélica dentro de indivíduos. Entretanto, subdividindo a população em várias subpopulações através de amostragem aleatória tem-se que a variância genética dentro de subpopulações equivalerá a $(1-F)$ vezes a variância genética original. E esta redução na variação genética é o que ocorre em populações com tamanho efetivo baixo (que são amostras de uma população base).

Por outro lado, a variação genética entre subpopulações aumenta, podendo conduzir a uma diferenciação entre elas. Maiores detalhes são descritos no item subdivisão e estrutura de populações.

21.4.5. Depressão endogâmica em populações silvestres e populações florestais

Em animais e plantas domésticas é sabida a existência de depressão endogâmica para o caráter adaptação (Falconer, 1989). Para populações silvestres, espera-se que a depressão endogâmica seja mais drástica do que em populações cativeiras devido à predominância de ambientes desfavoráveis, o que tem sido confirmado em vários estudos (Frankham, 1995a), ao contrário do que se pregava anteriormente.

A endogamia em populações silvestres ocorre como consequência dos efeitos do tamanho populacional finito ao longo de muitas gerações. Pode ocorrer, também, em populações de plantas devido ao cruzamento entre indivíduos parentados que tendem a se estabelecer proximamente (endogamia de vizinhança). Conforme Frankham (1995a), as estimativas de depressão endogâmica para caracteres componentes da adaptação (por exemplo, crescimento) subestimam a depressão para a adaptação propriamente dita, a qual é, via de regra, três vezes maior. Dessa forma, os reais impactos da depressão têm sido subestimados.

Em espécies florestais, tem sido observada grande depressão por endogamia em muitos estágios de vida, especialmente, em coníferas, as quais apresentam depressão superior àquela verificada para muitas plantas de ciclo curto, para *Drosophila* e para a espécie humana, conforme estimativas transformadas em números de equivalentes letais (Savolainen, 1994; Williams & Savolainen, 1996).

A depressão endogâmica pode ser minimizada sobretudo por meio da imigração, ou seja, do cruzamento entre diferentes populações de plantas. O outro fator que pode diminuir o impacto da depressão endogâmica é a eliminação de alelos deletérios por meio da endogamia prévia nas populações, seguida por seleção natural.

21.5. Polimorfismo e Variação Genética

A mutação é o processo fundamental na geração de variabilidade genética, pois provoca a geração de novos alelos na população, que levam ao aparecimento do polimorfismo genético nas populações.

O poliformismo genético pode ser definido como sendo a ocorrência de dois ou mais alelos na população em freqüências consideráveis, ou seja, a freqüência de cada um de pelo menos dois alelos deve ser superior a 1%. E, logicamente, este polimorfismo alélico implica também polimorfismo genotípico. Assim, o polimorfismo redonda na própria variabilidade genética. Locos com polimorfismo alélico são denominados locos polimórficos e aqueles sem polimorfismo alélico são chamados monomórficos. Os alelos (raros) com freqüência menor que 1% são denominados idiomorfos.

O polimorfismo genético (alélico e genotípico) pode ser avaliado principalmente por medidas de diversidade genética dentro de populações e medidas de identidade e distância genética entre populações. É importante relatar que aqui identidade significa alelos com efeitos indistinguíveis e não alelos idênticos por descendência (Hartl & Clark (1989). Na Tabela 21.2 são listados alguns dos principais estimadores de diversidade, identidade e distâncias genéticas segundo Hartl & Clark (1989) e Futuyma (1992).

Tabela 21.2. Medidas de similaridade e diversidade genética empregadas em estudos de genética de populações e seus respectivos estimadores.

Medidas	Estimadores*
Diversidade	
• Proporção de locos polimórficos (PLP)	PLP = NLP/NTL
• Número de alelos por loco polimórfico (NALP)	NALP = NTA/NLP
• Índice de diversidade de Nei ao nível de subpopulação (população) (H_S)	$H_S = 1 - \sum_{i=1}^m p_{is}^2$
• Índice de diversidade de Nei ao nível de população total (espécie) (H_T)	$H_T = 1 - \sum_{i=1}^m \bar{p}_i^2$
• Diversidade genética entre subpopulações (H_b)	$H_b = H_T - H_S$
• Índice de fixação de Wright ou redução de heterozigose (diversidade) na subpopulação ou divergência genética entre subpopulações	$F_{ST} = \frac{H_T - \bar{H}_S}{H_T}$
Identidade ou similaridade	
• Índice de similaridade genética de Nei (I_N)	$I_N = \frac{\sum_{i=1}^m (p_{ix} \cdot p_{iy})}{\left[\left(\sum_{i=1}^m (p_{ix}^2) \right) \left(\sum_{i=1}^m (p_{iy}^2) \right) \right]}$

Continua...

Tabela 21.2. Continuação.

Medidas	Estimadores*
• Índice de similaridade genética de Rogers (I_R)	$I_R = 1 - \left[(1/2) \sum_{i=1}^m (p_{ix} - p_{iy})^2 \right]^{1/2}$
• Identidade fenotípica de Hedrick (I_H)	$I_H = \frac{\sum_{g=1}^n (p_{gx} - p_{gy})}{(1/2) \left[\left(\sum_{g=1}^n p_{gx}^2 + \sum_{g=1}^n p_{gy}^2 \right) \right]}$
Distância genética	
• Distância genética de Nei	$D_N = -\log_e I_N$
• Distância Genética de Rogers	$D_R = 1 - I_R$
• Distância Genética de Hedrick	$D_H = 1 - I_H$

* NLP : número de locos polimórficos; NTL : número total de locos; NTA : número total de alelos nos locos polimórficos; p_i : freqüência do alelo i na subpopulação; \bar{p}_i : freqüência média do alelo i através das subpopulações; p_{ix} : freqüência do alelo i na subpopulação x ; p_{gx} : freqüência do genótipo g na população x ; m e n : números de alelos por loco e de genótipos.

O termo H na Tabela 21.1 refere-se à heterozigose que é uma medida de diversidade baseada em freqüências alélicas e não genotípicas, podendo assim ser aplicada a populações com qualquer sistema reprodutivo, sexuado ou assexuado. Para populações alógamas em equilíbrio de Hardy-Weinberg, H equivale à freqüência de heterozigotos. As freqüências alélicas e genotípicas necessárias aos estudos de diversidade e distância genética podem ser obtidas a partir da análise de marcadores genéticos ou bioquímicos. Estudos de diversidade genética em populações florestais usando marcadores têm sido comuns (Adams et al., 1992).

A análise da variabilidade genética via estudos de características quantitativas é extremamente útil na conservação genética, sobretudo em espécies nativas (Kageyama & Dias, 1985), fornecendo bases sólidas para inferências sobre a distribuição da variabilidade entre e dentro de populações. Para manutenção da variabilidade entre e dentro de populações, Namkoong (1997) afirma que é preciso trabalhar com pelo menos uma dezena de populações com tamanho efetivo adequado. Considerando que mais de 80% da diversidade genética total em espécies tropicais ocorre dentro de populações, Hamrick (1992) sugere que cinco populações, adequadamente amostradas, manteriam 99% da diversidade genética total das espécies.

Baixos níveis de variação genética, em geral, decorrentes de pequenos tamanhos efetivos populacionais e altos níveis de endogamia, tornam as populações vulneráveis aos fatores ambientais adversos, predispondo-as à extinção.

21.6. Mutação e Carga Genética

21.6.1. Carga genética

A mutação é um fenômeno raro e aleatório, geralmente ocasionando o aparecimento de alelos desfavoráveis e por isso recessivos. Isto gera nas populações (desde que os alelos não sejam neutros), o que se denomina de carga genética. Carga genética refere-se à redução do valor adaptativo da população ou indivíduo, como consequência da presença nas populações de alelos desfavoráveis gerados por mutação.

O conceito de carga genética foi criado por Haldane e uma série de fontes de carga podem ser enumeradas (Freire-Maia, 1974). A seguir serão descritas sucintamente algumas fontes de carga, mais relevantes em melhoramento genético.

Carga mutacional: composta por alelos deletérios ou desfavoráveis que estão sendo mantidos em equilíbrio nas populações.

Carga do intercruzamento: gerada pelo fluxo gênico (migração) de alelos desfavoráveis.

Carga de deriva genética: gerada quando a deriva reduz a freqüência dos alelos favoráveis.

21.6.2. Mutação e manutenção da variabilidade genética

Na natureza, as taxas de mutação são baixas, ocorrendo na faixa de 10^{-4} a 10^{-8} por geração, o que significa um em dez mil a um em 100 milhões de gametas. Assim, a mutação por si só produz mudanças muito lentas (gerando alelos idiomorfos) nas freqüências alélicas. E isto tende a ocorrer somente quando a mutação é recorrente, ou seja, para mutações que ocorrem repetidas vezes para o mesmo alelo.

A mutação recorrente previne a perda do alelo mutante por amostragem, já que contribui para aumentar a sua freqüência. A taxa de mutação recorrente é denominada de “pressão de mutação” e pode ser denominada de u , de forma que as freqüências alélicas para os genes selvagem B e mutante b passam a ser dadas por $p_o - \mu p_o$ e μp_o respectivamente, em que p_o é a freqüência inicial do alelo B.

Entretanto, existe uma taxa de mutação reversa (de b para B) que pode ser denominada v , de forma que a cada geração existe uma redução na freqüência de b dada por vq_o , em que q_o é a freqüência anterior de b. Assim, a mudança na freqüência alélica em uma geração é dada por: $\Delta q = \mu p_o - vq_o$.

O “equilíbrio ($\Delta q = 0$) de mutação” ocorre quando $\mu p - vq = 0$, ou seja para $q = \mu / (\mu + v)$.

Para a conservação genética, a mutação desempenha papel fundamental, principalmente, em dois aspectos: a geração e a manutenção da variabilidade genética.

A mutação vista como um fator mantenedor de variabilidade genética e creditada como um fator que não pode ser negligenciado pelos geneticistas de populações foi descrita apenas recentemente (Falconer, 1989).

A teoria de como a mutação afeta a seleção em caracteres quantitativos foi desenvolvida por Hill (1982). Variação é continuamente produzida por mutação em cada geração na taxa de $1/(2N)$, em que N é o tamanho da população. Com baixas freqüências, os alelos pouco contribuem para a variação genética, mas, por meio da seleção, as freqüências desses alelos podem ser aumentadas continuamente a ponto de contribuírem efetivamente para a variação genética e para a resposta à seleção. Entretanto, podem ser necessárias 20 gerações para que a mutação possa contribuir efetivamente para a seleção (evolução).

A taxa de ganho genético (evolução), devida à mutação, depende da intensidade de seleção (k), do tamanho efetivo populacional (N_e), do desvio padrão fenotípico (σ_F) e da variância genética aditiva resultante de novos mutantes em cada geração (σ_{am}^2), e o ganho genético por geração (na ausência de dominância) é dado por $G_s = 2N_e k \sigma_{am}^2 / \sigma_F$.

Assim, o ganho é maior em populações maiores, gastando, entretanto, mais tempo para atingir o valor máximo, pois, em grandes populações, os alelos mutantes iniciam-se em freqüências mais baixas.

Dessa forma, a mutação pode explicar a manutenção de variabilidade genética indefinidamente nas populações. Em princípio, o ganho genético advém de variabilidade genética aditiva presente na população. Posteriormente, a taxa de ganho diminui gradualmente devido à redução na variabilidade genética aditiva. Depois, após talvez 20 gerações, nova variação advinda de alelos mutantes começa a contribuir para o ganho genético, o qual passa a decrescer mais lentamente. Quando os alelos da população inicial estiverem fixados, o ganho genético dependerá apenas da variância mutacional e a taxa de ganho será constante indefinidamente. Em outras palavras, com mutação não ocorrerá limites seletivos (Falconer, 1989).

Estudos experimentais têm estimado σ_m^2 (variabilidade genética mutacional total) como aproximadamente $10^{-3} \sigma_E^2$, em que σ_E^2 é a variância ambiental (Falconer, 1989). Sendo $2N_e \sigma_m^2 + \sigma_E^2$ a variância fenotípica, a herdabilidade no sentido amplo associada à variabilidade mutacional equivale a

$$h_g^2 = \frac{2N_e \sigma_m^2}{2N_e \sigma_m^2 + \sigma_E^2} = \frac{0,002N_e}{1 + 0,002N_e} \quad (\text{considerando } \sigma_m^2 = 10^{-3} \sigma_E^2).$$

Assim, tem-se os seguintes valores para h_g^2 em função do tamanho populacional (Falconer, 1989).

$N_e = 10$	$N_e = 25$	$N_e = 50$	$N_e = 100$	$N_e = 500$	$N_e = 1.000$	$N_e = 10.000$
$h_g^2 = 0,02$	$h_g^2 = 0,05$	$h_g^2 = 0,09$	$h_g^2 = 0,17$	$h_g^2 = 0,50$	$h_g^2 = 0,67$	$h_g^2 = 0,95$

Conclui-se, a partir destes valores, que a mutação não é importante para a manutenção de variabilidade genética e do ganho com seleção, somente em populações extremamente pequenas.

21.6.3. Acúmulo e perda de mutações deletérias

Lande (1994) reporta que o acúmulo de mutações semideletérias em populações com tamanho efetivo maior que 1.000 pode ser a causa mais importante de extinção. Com relação à possível eliminação desses alelos deletérios por meio de endogamia forçada, Frankham (1995a) concluiu que esta prática pode apenas reduzir a depressão endogâmica, mas não eliminá-la.

21.7. Migração e Fluxo Gênico

A migração (entre populações) de alelos ou de indivíduos seguida por cruzamentos, caracteriza o que é denominado fluxo gênico ou alélico, o qual contribui para a alteração das freqüências alélicas.

Considerando uma população com freqüência alélica q_o que recebe uma fração m de imigrantes, advindos de uma outra população com freqüência alélica q_m , a freqüência alélica nesta população na próxima geração é dada por:

$q_1 = mq_m + (1 - m) q_o = m (q_m - q_o) + q_o$. A alteração (Δq) na freqüência alélica pode ser quantificada por $\Delta q = q_1 - q_o = m (q_m - q_o)$.

Assim, a alteração na freqüência alélica é dependente da taxa de migração e da diversidade alélica entre as populações envolvidas. Torna-se claro, então, que a migração só é efetiva quando, simultaneamente, a taxa de migração e a divergência alélica são consideráveis.

Em populações com base genética restrita, a migração efetiva pode atuar como um importante fator de ampliação de variabilidade genética e o seu conceito está sendo utilizado quando se formam compostos e populações híbridas para fins de conservação. Em conservação de germoplasma, a reunião de amostras de diferentes populações (Resende & Vencovsky, 1990) é também uma forma de se ampliar a variabilidade genética total de uma espécie, constituindo também uma forma de migração.

Em espécies arbóreas tropicais, o fluxo gênico em populações naturais é restrito, fato que demanda a amostragem de um maior número de indivíduos por sítio, do que em espécies temperadas (Hamrick, 1992).

No tópico a seguir, a migração é discutida no contexto da estrutura populacional.

21.8. Subdivisão e Estrutura de Populações

Populações que não se enquadram na definição teórica de panmíticas (grandes e sob cruzamentos aleatórios) são ditas populações estruturadas. A estruturação pode se dar através de subpopulações com limitada migração e fluxo gênico entre elas até subpopulações totalmente isoladas. Estas subpopulações são geralmente denominadas “isolados”. A subdivisão das populações provoca um excesso de homozigose ou seja um efeito de endogamia.

Wright (1951) introduziu os parâmetros (denominados “estatísticas **F** hierárquicas”) F_{ST} , F_{IT} e F_{IS} para descrever estrutura genética, onde:

F_{ST} : é a correlação de alelos entre indivíduos dentro de subpopulações. Refere-se, portanto, ao próprio **F** ou coeficiente de endocruzamento (de Wright) dentro das subpopulações, provocado pela própria deriva genética associada à subdivisão.

F_{IT} : é a correlação entre alelos dentro de indivíduos. Refere-se ao coeficiente de endocruzamento abrangendo toda a espécie (população).

F_{IS} : é a correlação entre alelos dentro de indivíduos dentro de subpopulações. Refere-se a um coeficiente de endocruzamento de um indivíduo devido a cruzamentos não aleatórios dentro de subpopulações.

$$F_{IS} = (F_{IT} - F_{ST}) / (1 - F_{ST})$$

O coeficiente F_{IT} resulta da endogamia (cruzamentos não aleatórios) dentro da subpopulação (F_{IS}) e da endogamia devida à subdivisão (F_{ST}). Se existir panmixia dentro de subpopulações, $F_{IS} = 0$ e, portanto, $F_{IT} = F_{ST}$. Se todas as populações forem geneticamente idênticas, $F_{ST} = 0$ e, portanto, $F_{IS} = F_{IT}$.

Sendo \bar{p} a freqüência alélica média sobre várias subpopulações e σ_p^2 a variância da freqüência alélica média, $F_{ST} = \sigma_p^2 / \bar{p}(1-\bar{p})$. Sendo $F_{ST} = 1/(2N_e)$, o N_e é estimado por meio de $\sigma_p^2 = \bar{p}(1-\bar{p}) / (2N_e)$.

Considerando que há equilíbrio nas subpopulações, $F_{ST} \equiv 1 / (1 + 4N_e m)$, em que m é a taxa de migração em cada subpopulação em cada geração e N_e é o tamanho efetivo com valor igual para todas as subpopulações. Tendo-se estimado F_{ST} , pode-se estimar $N_e m$ por $N_e m = (1 - F_{ST}) / (4F_{ST})$, o qual, assumindo valor <1, significa que a deriva conduzirá a substancial diferenciação. Esta abordagem de F_{ST} fornece um outro estimador para a variância das freqüências alélicas, ou seja, $\sigma_p^2 = \bar{p}(1-\bar{p}) / (1 + 4N_e m)$, mostrando que a variação entre subpopulações diminui à medida que $N_e m$ aumenta.

Os parâmetros mencionados podem ser estimados por meio da avaliação de marcadores genéticos codominantes obtidos da estrutura hierárquica: genes dentro de indivíduos (G), indivíduos dentro de progênies (I), progênies ou famílias (F) dentro de populações e populações (P) ou subpopulações. Neste esquema, a

variância da freqüência alélica pode ser estimada via análise de variância da variável aleatória discreta X, com o valor 1 para presença do alelo e 0 para ausência (Weir, 1990; Vencovsky, 1992). Trata-se, portanto, de uma distribuição Bernoulli com média p e variância $p(1-p)$.

O esquema da análise de variância das freqüências alélicas é apresentado a seguir:

Fontes de Variação	GL	QM	E(QM)
População (P)	(s-1)	Q_P	$\sigma_G^2 + 2\sigma_I^2 + 2n \sigma_F^2 + 2mn \sigma_P^2$
Famílias (F)/P	s(m-1)	Q_F	$\sigma_G^2 + 2\sigma_I^2 + 2n \sigma_F^2$
Indivíduos (I)/F	sm(n-1)	Q_I	$\sigma_G^2 + 2\sigma_I^2$
Genes (G)/I	smn	Q_G	σ_G^2
Total	2 smn-1		

Assim, a variância total σ_T^2 da freqüência do alelo em questão é dada por $\sigma_T^2 = p(1-p) = \sigma_P^2 + \sigma_F^2 + \sigma_I^2 + \sigma_G^2$. São verificadas as seguintes relações entre estes componentes de variância e os parâmetros descritivos da estrutura de populações:

$$\sigma_P^2 = p(1-p) F_{ST}$$

$$\sigma_F^2 = p(1-p)(\theta - F_{ST})$$

$$\sigma_I^2 = p(1-p)(F_{IT} - \theta)$$

$$\sigma_G^2 = p(1-p)(1-F_{IT})$$

em que θ é o coeficiente de coancestria entre os indivíduos nas progênieis.

Com equilíbrio de Wright (equilíbrio de endocruzamento) a taxa de autofecundação pode ser estimada a partir de $F_{IS} = S/(2-S) = (1-t)/(1+t)$ em que t é a taxa de cruzamento.

Assim, podem ser derivados os seguintes estimadores:

$$\hat{F}_{ST} = \hat{\sigma}_P^2 / \hat{\sigma}_T^2; \quad \hat{\theta} - \hat{F}_{ST} = \hat{\sigma}_F^2 / \hat{\sigma}_T^2;$$

$$\hat{F}_{IT} - \hat{\theta} = \hat{\sigma}_I^2 / \hat{\sigma}_T^2; \quad 1 - \hat{F}_{IT} = \hat{\sigma}_G^2 / \hat{\sigma}_T^2$$

$$\hat{F}_{IS} = \frac{\hat{F}_{IT} - \hat{F}_{ST}}{1 - \hat{F}_{ST}} = \frac{\hat{\sigma}_F^2 + \hat{\sigma}_I^2}{\hat{\sigma}_F^2 + \hat{\sigma}_I^2 + \hat{\sigma}_G^2};$$

$$\hat{S} = \frac{2\hat{F}_{IS}}{1 + \hat{F}_{IS}}; \quad \hat{t} = 1 - \hat{S}.$$

Os componentes de variância são obtidos via esperança de quadrados médios da análise de variância no esquema hierárquico, apresentado anteriormente. Um exemplo de cálculo destes componentes, foi apresentado, de maneira simples e clara, por Dias (1998b).

Pode-se observar que, em equilíbrio de F , $\theta = (1+S)^2/[4(2-S)]$, de forma que o coeficiente de coancestria pode ser estimado por meio da estimação de S .

O tamanho efetivo populacional associado ao total de indivíduos pode ser estimado por:

$$\sigma_p^2 = \frac{p(1-p)}{2N_e} = \sigma_p^2 / s + \sigma_F^2 / (ms) + \sigma_I^2 / (nms) + \sigma_G^2 / (2nms), \text{ em que } s, m \text{ e } n$$

referem-se ao número de populações, progênies e indivíduos, respectivamente.

Equivalentemente à probabilidade de que dois alelos tomados ao acaso de duas subpopulações sejam idênticos por descendência, o F_{ST} fornece também uma medida da variação ou diversidade genética entre as populações, provocada, sobretudo, por deriva genética, excluindo mutação e migração. F_{ST} representa a endogamia (homozigose) na subpopulação em relação à população total e é denominado “índice de fixação” de Wright (entre subpopulações), representando a redução na heterozigose da subpopulação devido à deriva genética aleatória.

Em espécies tropicais a subdivisão ou fragmentação representa um problema especial para a conservação da diversidade genética. A fragmentação de uma floresta inicialmente contínua, não conduzirá por si só à perda de variação genética. Uma vez que a maioria da diversidade está presente dentro de populações, fragmentos de tamanho moderado contendo de 15 a 20 indivíduos deverão conter a maioria da diversidade das espécies. Entretanto, estes fragmentos estarão sujeitos a novas derivas, e a variação genética diminuirá gradativamente com a perda destes fragmentos e, consequentemente, a diversidade geral da espécie será perdida (Hamrick, 1992). Este é um problema típico da erva-mate, por exemplo.

Se a subdivisão provoca aumento da homozigose, a quebra dos isolados, ou seja, a reunião das subpopulações provocará redução na freqüência média de homozigotos em relação à freqüência média antes da fusão. Este efeito é denominado “efeito ou princípio de Wahlund” e afeta apenas as freqüências genotípicas e não as freqüências alélicas. A seguir este efeito é descrito matematicamente, com base em Freire-Maia (1974).

Sendo X e Y dois isolados ou subpopulações de tamanhos iguais, as freqüências genotípicas são p^2 , $2pq$ e q^2 e r^2 , $2rs$ e s^2 para os genótipos BB, Bb e bb, respectivamente, para as subpopulações X e Y, respectivamente. p , q e r , s são as freqüências alélicas em X e Y, respectivamente.

Na população formada pelo conjunto das subpopulações (ainda isoladas), as freqüências alélicas médias são $(p+r)/2$ para B e $(q+s)/2$ para b e as freqüências genotípicas médias são $(p^2 + r^2)/2$, $(2pq + 2rs)/2$ e $(q^2 + s^2)/2$ para BB, Bb e bb, respectivamente.

Com a fusão, as freqüências alélicas continuam as mesmas, mas as freqüências genotípicas serão $[(p+r)/2]^2$, $2[(p+r)/2](q+s)/2$ e $[(q+s)/2]^2$ para BB, Bb e bb, respectivamente, num total de $[(p+r)/2 + (q+s)/2]^2$. O decréscimo na freqüência dos homozigotos é dado por $(p^2 + r^2)/2 - [(p+r)/2]^2$. Esta redução equivale exatamente à variação da freqüência do alelo B entre as subpopulações, que, de modo genérico equivale a

$$\sigma_p^2 = \sum_{i=1}^n p_i^2 / n - \left[\frac{\sum_{i=1}^n p_i}{n} \right]^2, \text{ em que } n \text{ é o número de subpopulações.}$$

Em espécies tropicais, o fluxo gênico é essencial na manutenção dos fragmentos por meio da própria quebra dos isolados. Conforme Hamrick (1992), se o fluxo gênico ocorre a distâncias de 2 km, a diversidade genética dentro de fragmentos pode ser mantida, e consequentemente, poucos fragmentos dentro de uma região geográfica serão necessários para preservar a diversidade da espécie.

21.9. Estimação da Herdabilidade em Populações Naturais Empregando Dados Morfológicos e Marcadores Genéticos

Para a estimação de parâmetros genéticos tais como a herdabilidade, são necessárias informações fenotípicas e de parentesco entre os indivíduos avaliados. Uma vez que as análises genéticas de dados moleculares fornecem informações sobre o parentesco entre os indivíduos, torna-se possível a estimativa de parâmetros genéticos a partir da avaliação fenotípica dos indivíduos nas próprias condições naturais. Ritland (1996) apresenta estimadores para a herdabilidade nestas condições.

Definindo $Z_i = \frac{(y_i - \bar{y})(y_i' - \bar{y})}{Var(y)}$, como a similaridade fenotípica entre dois indivíduos na população, em que y_i e y_i' referem-se a observações fenotípicas nos indivíduos i e i' , \bar{y} e $Var(y)$ são relativos à média e à variância do caráter y na

população, tem-se: $Z_i = \frac{2r_i \hat{\sigma}_a^2 + e_i}{\sigma_y^2} = 2r_i h_a^2 + \frac{e_i}{\sigma_y^2}$, em que:

r_i : coeficiente de parentesco de Malecot.

$h_a^2 = \frac{\sigma_a^2}{\sigma_y^2}$: herdabilidade individual no sentido restrito.

$e_i = \frac{e_i}{\sigma_y^2}$: resíduo devido aos efeitos ambientais.

σ_a^2 : variância genética aditiva.

$\sigma_y^2 = Var(y)$.

O estimador da herdabilidade é dado por $\hat{h}^2 = cov(Z, r) / [2 Var(r)]$, pois $h^2 = \frac{cov(Z, r)}{2 Var(r)} = \frac{cov[(2r h^2 + e_i'), r]}{2 Var(r)} = \frac{2 Var(r) h^2}{2 Var(r)} = h^2$

Assim, de posse dos coeficientes de similaridade fenotípica (Z) e dos coeficientes de parentesco de Malecot (r), estima-se facilmente a herdabilidade como um coeficiente de regressão \hat{h}^2 que envolve as variáveis Z e r .

21.10. Procedimentos *Bootstrap* e *Jackknife* no Estudo de Parâmetros em Genética de Populações

Os métodos de *Bootstrap* e *Jackknife* são processos computacionalmente intensivos de reamostragem, no próprio conjunto original de dados, em que amostras sucessivas são retiradas deste conjunto de dados. Em cada amostra são obtidas as estimativas dos parâmetros de interesse, empregando os mesmos estimadores usados no conjunto de dados originais. A variância entre as estimativas obtidas nas várias amostragens são utilizadas para caracterizar a distribuição dos estimadores (estimação da distribuição amostral das estimativas), para a obtenção de estimativas dos desvios padrões das estimativas e construção de intervalos de confiança. São muito úteis no estudo de estatísticas com distribuição desconhecida, ocasião esta em que os procedimentos analíticos não podem ser utilizados.

Bootstrap

O *bootstrap* (Efron, 1979; 1982; 1987; Efron & Tibshirani, 1986) é uma técnica estatística relacionada ao *jackknife* (esta foi desenvolvida primeiro e é apresentada no tópico seguinte) e foi desenvolvida por Efron (1979) (o nome *bootstrap* pode significar em vantagem ou além da armadilha ou cilada ou golpe). Baseia-se em retiradas de amostras aleatórias de tamanho n , com reposição, a partir do conjunto de dados observados. Assim, a amostragem é composta, a cada passo, sempre com o número total de dados, de forma que cada observação tem sempre a mesma probabilidade de ser amostrada, em cada passo. Esta amostragem com reposição implica que a amostra *bootstrap* pode conter observações duplicadas e observações faltantes, em relação ao conjunto original de dados.

A partir da amostra *bootstrap*, obtêm-se as estimativas dos parâmetros (θ) de interesse, empregando-se os estimadores correntes. Estas são as primeiras estimativas *bootstrap* $\hat{\theta}^*$. O processo de reamostragem com reposição é repetido um grande número de vezes, obtendo-se várias estimativas *bootstrap* $\hat{\theta}^*$. A partir de todas as estimativas $\hat{\theta}^*$, computa-se a variância de $\hat{\theta}^*$, a qual é uma estimativa da variância da distribuição de amostragem de $\hat{\theta}$ e, portanto, permite inferir sobre a precisão da estimativa $\hat{\theta}$, obtida com o conjunto original de dados ou como a média das estimativas *bootstrap*.

A lógica inerente ao método *bootstrap* refere-se ao fato de que, ao tomar 95% dos dados originais, se os resultados mostrarem-se completamente diferentes dos obtidos com o conjunto total de dados, não se pode ter muita confiança neles. Isto porque poder-se-ia ter, na amostra original, apenas 95% dos

dados. Assim, o *bootstrap* revela a consistência interna dos dados, ou seja, se as estimativas mudam muito com a reamostragem dos dados, deve-se ter menor confiança em tais estimativas.

Vencovsky et al. (1997) resumem os passos associados à aplicação do *bootstrap*. Sendo θ o parâmetro de interesse e $\hat{\theta}$, o seu estimador obtido a partir de um conjunto de dados (Y_1, Y_2, \dots, Y_n), o procedimento consiste em:

- (a) Aproximar a distribuição desconhecida F da estatística de interesse, por uma distribuição discreta \hat{F} , com pesos $1/n$ para cada ponto da amostra original de dados.
- (b) Retirar de \hat{F} uma amostra de tamanho n , com reposição, denominada amostra *bootstrap* e calcular a estatística *bootstrap* $\hat{\theta}^*$.
- (c) Repetir o passo (b), independentemente, b vezes (em geral, $50 \leq b \leq 200$).
- (d) Obter a estimativa *bootstrap* do parâmetro θ dada por $\hat{\theta}_m^* = \frac{1}{b} \sum_{i=1}^b \hat{\theta}_i^*$.
- (e) Calcular o vício *bootstrap* por $\hat{V} = \hat{\theta} - \hat{\theta}_m^*$.
- (f) Calcular a variância *bootstrap* por $\sigma_{\hat{\theta}^*}^2 = \frac{1}{b-1} \sum_{i=1}^b (\hat{\theta}_i^* - \hat{\theta}_m^*)^2$.
- (g) Construir intervalos de confiança para θ .

Os intervalos de confiança para θ podem ser obtidos de três maneiras diferentes: (i) pelo método dos percentis; (ii) pelo método *bootstrap* corrigido para o vício; (iii) pelo método *t* de Student, em que se considera $\frac{\hat{\theta}_m^* - \theta}{\hat{\sigma}_{\hat{\theta}^*}} \sim t_{n-1}$. Por este último método, os limites inferior e superior do intervalo de confiança são dados por $LI = \hat{\theta}_m^* - t_{(n-1, \frac{\alpha}{2})} \hat{\sigma}_{\hat{\theta}^*}$ e $LS = \hat{\theta}_m^* + t_{(n-1, 1-\frac{\alpha}{2})} \hat{\sigma}_{\hat{\theta}^*}$.

Jackknife

O *jackknife* (Miller, 1974; Lebart et al., 1979; Efron, 1982) é uma técnica de reamostragem na qual é usada apenas uma fração dos dados originais, tomando-se amostras repetidas sem reposição. A técnica foi originalmente proposta por Quenouille (1956) e sofreu modificações por Tukey (1958) e tem sido aplicada com sucesso em estatísticas de diversidade biológica (Zahl, 1977; Weir & Cockerham, 1984). O nome da técnica foi escolhido por Tukey, na esperança de que se trataria de uma robusta e pronta ferramenta estatística (a tradução de *jackknife* equivale a algo como um canivete pontiagudo, de bolso).

O *jackknife* é um procedimento que permite melhorar a estimativa de qualquer parâmetro. Apresenta grande flexibilidade de uso, pois não se baseia em qualquer suposição quanto à distribuição dos dados. O método baseia-se em estimativas *jackknife* as quais são normalmente distribuídas e cuja média

fornecer a melhor estimativa do parâmetro. Intervalos de confiança podem ser construídos em associação com esta melhor estimativa. Esta técnica apresenta grande utilidade na estimação de medidas e índices de diversidade genética, os quais, muitas vezes, não apresentam distribuição amostral conhecida.

O procedimento *jackknife* é resumido a seguir. Na estimação de um parâmetro θ a partir de uma amostra ou conjunto de dados com n observações, o procedimento *jackknife* para a estimação da variância do estimador $\hat{\theta}$ consiste na omissão de cada uma das n observações, uma em cada reamostragem, e no uso da variação entre as n estimativas resultantes. Sendo $\hat{\theta}_i$ a estimativa do parâmetro obtida quando da omissão da observação i , tem-se que:

$$\hat{\theta}^* = n\hat{\theta} - \frac{n-1}{n} \sum_{i=1}^n \hat{\theta}_i : \text{estimador melhorado (menos viciado) de } \theta;$$

$$\hat{\sigma}_{\hat{\theta}}^2 = \frac{n-1}{n} \sum_{i=1}^n \left(\hat{\theta}_i - \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i \right)^2 : \text{variância jackknife de } \hat{\theta}.$$

O método tem sido muito utilizado na estimação da variância de estimadores de proporções, como a herdabilidade. A metodologia tradicional padrão usa a distribuição qui-quadrado para a construção de intervalos de confiança para variâncias, a partir da variância amostral. De maneira similar, a distribuição F de Snedecor é utilizada para inferências sobre uma razão entre variâncias, a partir da razão entre variâncias amostrais independentes. A teoria distribucional destas técnicas é precisamente correta apenas quando as observações são normalmente distribuídas. A técnica *jackknife* tem-se mostrado capaz de estimar corretamente a variabilidade de estimativas de parâmetros, para distribuições mais gerais (Miller, 1974).

A metodologia generalizada do *jackknife* baseia-se na divisão do conjunto de N dados amostrais em g grupos de tamanho igual a k , de forma que $N = gk$. Em geral, k é tomado como 1, mas, pode ser tão grande quanto $N/2$. O estimador $\hat{\theta}_i$ corresponde àquele baseado em amostras de tamanho $(g - 1)k$, onde o i -ésimo grupo de tamanho k foi removido. Com $k = 1$, $N = g$ e $(g - 1)k = g - 1$, de forma que $\hat{\theta}_i$ refere-se à amostra em que foi omitida a observação i .

21.11. Seleção

Em conservação genética, duas modalidades de seleção podem ser consideradas associadas à reprodução nas passagens de geração: a seleção massal ou individual e a seleção dentro de famílias (Nagylaki, 1992).

A seleção massal atua no conjunto total de descendentes e pode conduzir à seleção (eliminação) de famílias. Já a seleção dentro de famílias, mantendo-se igual número de indivíduos/famílias, elimina o componente de seleção entre famílias e praticamente duplica o tamanho efetivo populacional. Embora a sele-

ção dentro de famílias seja favorável por aumentar o N_e , ela conduz também a um aumento da carga de mutação em decorrência do aumento do N_e . Entretanto, provavelmente seja mais importante e praticável maximizar o N_e do que reduzir a carga genética.

Na conservação genética ex situ associada aos testes de progênies e programas de melhoramento, outras modalidades de seleção podem ser enunciadas como a seleção pelo BLUP (usando informação do indivíduo e da família) e a seleção entre famílias, as quais tendem a reduzir drasticamente o N_e . Nesta situação, os métodos mais favoráveis em termos de conservação da variabilidade genética são pela ordem: seleção dentro de famílias, seleção massal, seleção pelo BLUP e seleção entre famílias. Via de regra estes métodos conduzem aos menores ganhos genéticos obedecendo à mesma ordem, exceto a seleção entre famílias.

21.12. Amostragem para Estudos em Genética de Populações

O advento dos marcadores moleculares (DNA e isoenzimas) conduziu à possibilidade de obtenção de estimativas mais precisas de parâmetros em estudos de genética de populações. Entretanto, para que esse potencial aumento de precisão se concretize, torna-se necessária a utilização de processos adequados de amostragem, os quais, em essência, permitam preencher o requisito de representatividade das amostras.

Em geral, tem-se percebido, nos trabalhos publicados e/ou submetidos à publicação, grande ênfase nos aspectos metodológicos laboratoriais e moleculares, em detrimento dos aspectos de amostragem. Muitas vezes, a amostragem é deficiente, fato que compromete todo o trabalho. Em várias oportunidades, o processo de amostragem e o tamanho da amostra não são sequer mencionados. Assim, cabe ao leitor, fazer uma triagem dos trabalhos com resultados mais precisos, tendo em vista a amostragem realizada. Logicamente, os trabalhos que não abordam o tamanho amostral empregado e nem os desvios padrões das estimativas, devem ser descartados pelo leitor.

Entretanto, o ideal é que os próprios autores se conscientizem da importância da questão e, a partir disto, possam gerar resultados mais fidedignos. A seguir, são apresentadas algumas linhas gerais que podem nortear o planejamento de uma amostragem adequada. É importante destacar que um trabalho baseado em refinadas técnicas moleculares, mas com amostragem insuficiente, conduz à eficiência prática muito inferior àquela obtida por meio de técnicas convencionais de campo.

Amostragem dentro de indivíduos

A amostragem dentro de indivíduos diz respeito, essencialmente, à amostragem do genoma. Para dados fenotípicos ou morfológicos, a amostragem do genoma está diretamente relacionada ao número de caracteres avaliados, sendo que para cada caráter quantitativo são amostrados indiretamente vários locos.

Quando são utilizados marcadores genéticos baseados em isoenzimas, a amostragem do genoma depende do número de sistemas isoenzimáticos empregados (número de locos) e do número de alelos por loco. Neste caso, o número de sistemas a ser empregado varia com os objetivos. Por exemplo, para estimativas de coeficientes de parentesco entre indivíduos de populações naturais, Ritland (1996) recomenda que se deve adotar $25 < n(m - 1) < 100$, em que n refere-se ao número de locos e m equivale ao número de alelos por loco. Neste caso, 15 locos com 3 alelos ou 3 locos com 10 alelos seriam adequados, por exemplo.

No caso da utilização de marcadores do tipo RAPD, a amostragem dentro de indivíduos é menos problemática. Via de regra, tal procedimento propicia uma boa cobertura do genoma, visto que o DNA total é utilizado.

Amostragem dentro de progênies

Neste caso, uma abordagem baseada no conceito do tamanho efetivo populacional (N_e) fornece uma medida adequada da representatividade de uma progénie.

Para espécies alógamas, a expressão do tamanho efetivo de uma progénie (N_{ef}) de meios-irmãos (ou de polinização aberta) é dada por:

$$N_{ef} = \frac{4n}{n+3}, \text{ em que } n \text{ é o número de indivíduos por progénie.}$$

Baseando-se nesta expressão, pode-se inferir sobre a amostragem dentro de progênies necessária para atingir uma fração da representatividade completa ($n = \infty$ indivíduos) da progénie, conforme a Tabela 21.3.

Tabela 21.3. Tamanho efetivo de uma progénie de meios-irmãos (N_{ef}) e fração do tamanho efetivo máximo de uma progénie (N_{efmax}), em função do número (n) de indivíduos amostrados por progénie.

<i>n</i>	<i>N_{ef}</i>	<i>N_{efmax}</i>	Fração do <i>N_{efmax}</i>
1	1	-	0,250
5	2,5	-	0,625
7	2,8	-	0,700
10	3,1	-	0,775
12	3,2	-	0,800
15	3,3	-	0,825
18	3,4	-	0,850
20	3,5	-	0,875
25	3,6	-	0,90
30	3,64	-	0,91
50	3,77	-	0,94
100	3,88	-	0,97
∞	4,00	4,00	1,00

Verifica-se que um número entre 10 e 25 indivíduos por progênies já é suficiente para representá-la adequadamente. Com $n > 25$ indivíduos por progênies, os acréscimos na representatividade são muito lentos (Tabela 21.3).

Amostragem dentro de populações

Neste caso, a representatividade genética de uma população depende do número de progêneras amostradas e do número de indivíduos amostrados por progênies. Tal representatividade pode ser aferida tendo por base as freqüências dos alelos em amostras de diferentes tamanhos efetivos.

Para espécies alógamas, as expressões do tamanho efetivo, em função do número de progêneras (N_f) e número (n) de indivíduos amostrados por progênies são dadas por:

$$N_e = \frac{4N_f n}{n+3}, \text{ para espécies monóicas.}$$

$$N_e = \frac{4N_f n}{n+1}, \text{ (aproximadamente) para espécies dióicas.}$$

Baseando-se na expressão para espécies monóicas, várias inferências podem ser realizadas com referência à amostragem genética em populações (Tabela 21.4).

Tabela 21.4. Amostragem genética em populações em função do número total de indivíduos (N), número de progêneras (N_f) e número (n) de indivíduos por família amostrados, quantificada por meio do tamanho efetivo populacional (N_e), freqüência mínima dos alelos retidos (FAR) e endogamia potencial ($F = 1/(2N_e)$).

N	N_f	n	N_e	FAR	F
100	100	1	100	0,03	0,005
100	50	2	80	0,04	0,006
100	25	4	57	0,06	0,009
100	20	5	50	0,06	0,010
100	10	10	31	0,08	0,016
80	80	1	80	0,04	0,006
30	30	1	30	0,08	0,017
10	10	1	10	0,17	0,050
200	20	10	62	0,04	0,008
500	20	25	71	0,04	0,007
250	25	10	77	0,04	0,006
500	50	10	154	0,02	0,003

A freqüência dos alelos retidos em uma determinada amostra pode ser quantificada com base no conceito de variância das freqüências alélicas entre várias amostras. A estimativa da variância das freqüências alélicas (p_1) entre várias amostras é dada por $\hat{\sigma}_{p_1}^2 = [p_0(1-p_0)]/[2N_e]$, em que:

$\sigma_{p_1}^2$: variância da freqüência de um alelo, entre amostras de tamanho $2N_e$.

p_0 : freqüência paramétrica do alelo na população original.

Com base nesta variância de amostragem binomial, pode-se construir um intervalo de confiança (I.C.) para a freqüência alélica em uma dada amostra, por meio da expressão:

$$I.C. = p_0 \pm z \{ [p_0(1-p_0)]/[2N_e] \}^{1/2} = p_0 \pm 1,96 \{ [p_0(1-p_0)]/[2N_e] \}^{1/2}, \text{ em que:}$$

z : valor tabelado da distribuição normal padrão associado a determinado grau de confiança, equivalendo a 1,96 para um intervalo construído com 95% de confiança.

A partir desta expressão e de posse da estimativa do N_e , pode-se inferir sobre que tipos (em termos de suas freqüências na população original) de alelos foram capturados pela amostra. Como exemplo, considere um alelo com freqüência 0,10 na população-base. Tomando uma amostra com dez progênies desta população, com um indivíduo de cada uma, tem-se $N_e = 10$. Neste caso, o intervalo de confiança para a freqüência do referido alelo na amostra é dado por:

$$I.C. = 0,10 \pm 1,96 \{ [0,10(0,90)]/[2(10)] \}^{1/2} = 0,10 \pm 1,96 (0,06708) = -0,03 \text{ a } 0,2314.$$

Verifica-se que o limite inferior do intervalo de confiança atingiu o valor zero. Assim, pode-se inferir que a amostra não conseguiu capturar alelos com freqüência igual ou inferior a 10% na população. Tomando-se agora um alelo com freqüência 0,17 na população base e considerando a mesma estratégia de amostragem, o intervalo de confiança da freqüência alélica é dado por:

$$I.C. = 0,17 \pm 1,96 \{ [0,17(0,83)]/[2(10)] \}^{1/2} = 0,17 \pm 1,96 (0,08399) = 0,0053 \text{ a } 0,3346.$$

Neste caso, o limite inferior do intervalo de confiança da freqüência alélica não atingiu o valor zero. Então, conclui-se que a amostra com $N_e = 10$ conseguiu capturar alelos com freqüência maior ou igual a 17% na população.

Com base nos valores de FAR da Tabela 22.4, pode-se inferir que uma amostra com $N_e = 10$ é totalmente inadequada, em termos de representatividade da população. Amostras com $N_e \geq 60$ já conseguem capturar alelos com freqüência $\geq 4\%$, o que é adequado. Para atingir este N_e de 60, várias opções existem, desde 60 progênies com um indivíduo por progénie até 15 progênies com 100 indivíduos por progénie (Tabela 21.5).

Tabela 21.5. Estratégias para conseguir uma amostragem adequada ($N_e = 60$) de uma população, considerando várias alternativas, em termos do número total de indivíduos (N), número de progênies de meios-irmãos (N_f) e do número (n) de indivíduos amostrados por progênie.

N	N_f	n	N_e
60	60	1	60
76	38	2	61
90	30	3	60
108	27	4	62
120	24	5	60
200	20	10	62
425	17	25	61
1500	15	100	58

Verifica-se que uma amostragem adequada de uma população envolve, no mínimo, 15 progênies (com muitos indivíduos de cada uma) ou 60 indivíduos (com um indivíduo por progênio). Para manter coerência (quando se deseja realizar inferências também sobre progênies) com o número de indivíduos necessário para representar uma progênio, a amostragem mínima adequada por população seria de 200 indivíduos pertencentes a 20 progênies (Tabela 21.5).

Outra abordagem para determinação do tamanho amostral ótimo para comparação entre populações baseia-se na acurácia requerida para uma precisão avaliação do valor genotípico da população. Neste caso, a acurácia é dada por (Resende, 1999b):

$$r_{gg} = \left[\frac{[n N_f \rho_p + (1 - \rho_a)(1 - \rho_p) + n (1 - \rho_p) \rho_a] h^2}{1 + (n-1)[\rho_a(1 - \rho_p) + \rho_p] h^2 + (N_f - 1) n \rho_p h^2} \right]^{1/2}, \text{em que:}$$

h^2 : herdabilidade individual total (entre + dentro de populações) do caráter

ρ_a : correlação genética aditiva entre os indivíduos da progênio ($\rho_a = 0,25$ para progênies de meios-irmãos).

ρ_p : correlação genética entre indivíduos de uma mesma população, mas de diferentes progênies ($\rho_p = \sigma_p^2 / \sigma_T^2$, ou seja, equivale à relação entre as variâncias genética entre populações e genética total).

Considerando-se que em espécies tropicais, cerca de 80% da variação genética total das espécies encontra-se dentro de populações (Hamrick, 1992), pode-se considerar $\rho_p = 0,20$. Substituindo-se os coeficientes de correlação genética entre indivíduos, a expressão da acurácia simplifica-se para:

$$r_{\hat{g}g} = \left[\frac{[0,20 n N_f + 0,60 + 0,20 n] h^2}{1 + (n-1) 0,40 h^2 + (N_f - 1) 0,20 n h^2} \right]^{1/2}$$

Escolhendo-se a priori, valores de acurácia de 90% ou 95%, pode-se determinar N_f e n (Tabela 21.6).

Tabela 21.6. Valores de acurácia ($r_{\hat{g}g}$) na predição de valores genotípicos de populações, em função da variação na herdabilidade (h^2), no número de progêneres (N_f), número de indivíduos por progênio (n) e porcentagem da variação genética entre populações (ρ_p), adotando-se $N = 180$ indivíduos por população.

$r_{\hat{g}g}$	h^2	ρ_p	N_f	n	N
0,90	0,10	0,20	10	18	180
0,90	0,10	0,20	15	12	180
0,90	0,10	0,20	20	9	180
0,90	0,10	0,20	25	7	180
0,95	0,20	0,20	10	18	180
0,95	0,20	0,20	15	12	180
0,95	0,20	0,20	20	9	180
0,95	0,20	0,20	25	7	180
0,93	0,10	0,30	10	18	180
0,93	0,10	0,30	15	12	180
0,93	0,10	0,30	20	9	180
0,93	0,10	0,30	25	7	180

Verifica-se que os resultados (Tabela 21.6) são aproximadamente coincidentes com aqueles obtidos pela abordagem do N_e . Adotando-se 180 indivíduos por população, atingem-se acurárias de 90% quando a herdabilidade é 0,10 e 95% quando a herdabilidade é 0,20. Constatata-se, também, que é mais importante o número total de indivíduos do que a variação no número de progêneres. Variando a porcentagem (ρ_p) da variação genética entre populações, obtém-se maior acurácia quando esta porcentagem é maior e menor acurácia quando a porcentagem é menor. Apenas quando o ρ_p for bem abaixo de 0,20, deve-se aumentar a amostragem por população.

Uma terceira abordagem refere-se ao uso do intervalo de confiança para a média amostral (\bar{y}), dada por: $I.C. = \bar{y} \pm z s(\bar{y}) = \bar{y} \pm 1,96 s(\bar{y})$, em que $s(\bar{y}) = [\hat{\sigma}^2 / N]^{1/2}$ é o erro padrão da média. Neste caso, pode-se estabelecer um erro tolerável (δ) na estimativa da média, dado por $\delta = 1,96 s(\bar{y}) = 1,96 [\hat{\sigma}^2 / N]^{1/2}$.

A partir desta expressão, obtém-se $N = \frac{(1,96)^2 \hat{\sigma}^2}{\delta^2}$, como o tamanho adequado da amostra, para um erro tolerável δ , escolhido a priori.

Assim, para determinação do tamanho amostral N, necessita-se estabelecer um erro tolerável (δ) na estimativa da média e ter uma estimativa $\hat{\sigma}^2$ da variabilidade fenotípica na população. O erro δ pode ser especificado em porcentagem da média, por exemplo, 10%. Neste caso δ é dado por $0,10 \bar{y}$ e $N = \frac{(1,96)^2 \hat{\sigma}^2}{(0,10 \bar{y})^2} = \frac{(1,96)^2 \hat{\sigma}^2}{(0,10)^2 \bar{y}^2} = \frac{(1,96)^2}{(0,10)^2} (CV)^2$, em que $CV = \frac{\hat{\sigma}}{\bar{y}}$ é o coeficiente de variação do caráter na população.

Assim, por esta abordagem, torna-se necessária apenas uma estimativa ou conhecimento prévio do coeficiente de variação fenotípico individual na população e quanto maior este maior será o tamanho necessário da amostra. Considerando desejável um erro de 10%, ter-se-iam os seguintes tamanhos adequados da amostra em função de diferentes coeficientes de variação na população, conforme a Tabela 21.7.

Tabela 21.7. Amostragem adequada por população, quantificada pelo número total de indivíduos (N), em função de alguns coeficientes de variação fenotípico individual na população.

CV	N	CV	N
0,10	4	0,50	96
0,20	15	0,60	138
0,30	35	0,70	188
0,40	61	0,80	246

Para variáveis binomiais e usando-se a aproximação à distribuição normal, a mesma expressão para N pode ser utilizada, bastando substituir $\hat{\sigma}^2$ por pq e \bar{y} por p, em que p e q referem-se às proporções observadas das duas classes fenotípicas. Neste caso, torna-se necessária apenas uma estimativa prévia de p, a proporção média observada, visto que $q = 1 - p$.

Amostragem dentro de espécies

Tendo em vista que mais de 80% da diversidade genética total em espécies tropicais encontra-se dentro de populações, Hamrick (1992) sugere que 5 populações, adequadamente amostradas, deverão contemplar 99% da diversidade total das espécies. Assim, uma amostragem adequada de uma espécie, envolveria entre 1.000 (5 populações, 20 progênies por população e 10 indivíduos por progénie) a 300 indivíduos (5 populações, 60 progênies por população e 1 indivíduo por progénie). Entretanto, para a estimação de componentes de variância (herdabilidade, por exemplo) com precisão, amostragens maiores são necessárias (geralmente mais que 1.000 indivíduos), conforme mostrado no Capítulo 14.

21.13. Estratégias para Implantação de Bancos de Germoplasma de Plantas Perenes

Os bancos de germoplasma são muito numerosos e apresentam grande relevância para os programas de melhoramento de espécies perenes, pois, em geral, funcionam ao mesmo tempo como banco genético e como teste de introdução de espécies, populações e clones em diferentes regiões. Dessa forma, tais bancos, muitas vezes, fornecem materiais diretamente para os plantios produtivos, sem passar por testes experimentais adicionais.

Nestes bancos de germoplasma, geralmente cada acesso encontra-se representado por uma a dez plantas, obtidas por via sexuada ou assexuada, plantadas em uma única fileira. Neste caso, se os dados produtivos obtidos nos bancos de germoplasma são utilizados para inferências sobre as produtividades comparativas dos acessos, tem-se uma comparação viciada entre tais acessos visto que os princípios experimentais de repetição e casualização não foram observados. Recomenda-se, por ocasião da instalação de um banco, distribuir as várias plantas de um acesso em diferentes repetições casualizadas, de forma que os dados produtivos obtidos possam ser usados efetivamente na caracterização, avaliação, comparação e seleção de acessos em bancos de germoplasma.

Nos casos em que as plantas nos bancos de germoplasma ou em plantios de produtores não se encontram identificadas por matriz ou clone de origem, a única forma de proceder a uma seleção mais acurada é por meio da realização de uma estratificação ambiental (pela divisão da área de plantio em vários blocos, a posteriori) e da adoção de medidas repetidas em cada uma das plantas, praticando-se a “seleção massal estratificada baseada em medidas repetidas ou em médias de safras”.

Quanto ao tamanho amostral, as diretrizes apresentadas no tópico anterior podem ser seguidas. No entanto, é importante ressaltar a representatividade de um acesso. Se um acesso refere-se a um clone, apenas um representante no banco de germoplasma é adequado para mantê-lo integralmente, desde que se garanta a sua sobrevivência e que ele seja utilizado clonalmente.

Se o acesso refere-se a uma progénie de polinização aberta de população alógama, para mantê-lo, é necessário incluí-lo no banco com pelos 10 indivíduos. Neste caso, o tamanho efetivo do acesso é três e não um. Se cada progénie é representada apenas por um indivíduo no banco, o acesso passa a ser um clone, com tamanho efetivo um.

21.14. Recomendações

As seguintes estratégias e ações podem ser adotadas visando à conservação genética de espécies predominantemente alógamas:

- a) Trabalhar com populações mais variáveis geneticamente.

- b) Maximizar o intervalo entre gerações.
- c) Maximizar a relação N_e/N .
- d) Manter igual tamanho de família.
- e) Manter a proporção de sexo próxima de 1:1.
- f) Manter um valor de N_e desejável e constante através das gerações.
- g) Utilizar parcelas com poucas plantas e muitas repetições;
- h) Adotar N_e na faixa de 500-5000 para conservação in situ e N_e de 150 (40 famílias) por população para conservação ex situ.
- i) Trabalhar com pelo menos cinco a dez populações.
- j) Praticar intercruzamento entre populações em gerações avançadas de conservação.
- k) Para a conservação associada a programas de melhoramento, dar preferência aos métodos de seleção dentro de progênies e seleção massal, em detrimento da seleção pelo BLUP e entre famílias.
- l) Para estudos em genética de populações, tamanhos amostrais da ordem de 180 a 200 indivíduos por população, pertencentes a cerca de 20 progênies, são adequados. Menor número de indivíduos pode ser empregado, desde que associado a maior número de progênies.

Capítulo 22

Referências Bibliográficas



- ABDEL-AZIM, G. A.; BERGER, P. J. Properties of threshold model predictions. *Journal of Animal Science*, v. 77, p. 582-590, 1999.
- ABE, K.; SATO, P.; SAITO, T.; KURIHARA, A.; KOTOBUTI, K. Inheritance of ripening time of fruit of japanese pear. *Japanese Journal of Breeding*, v. 43, n. 2, p. 289, 1993.
- ABE, K.; SATO, P.; SAITO, T.; KURIHARA, A.; KOTOBUTI, K. Narrow-sense heritability of fruit characters in japanese pear. *Breeding Science*, v. 45, n. 1, p. 1-5, 1995.
- ADAMS, W. T.; STRAUSS, S. H.; COPES, D. L.; GRIFFIN, A. R. *Population genetics of forest trees*. Kluwer: Dordrecht, 1992. 420 p.
- AITKEN, A. C. A note on selection from a multivariate normal population. *Proceedings of the Royal Society of Mathematics*, Edinburgh, Section A, v. 4, p. 106-110, 1934.
- AKAIKE, H. A new look at the statistical model identification. *IEEE Trans. Automat. Control*, v. 19, p. 716-723, 1974.
- ALLAIRE, F. R. Mate selection by selection index theory. *Theoretical and Applied Genetics*, v. 57, p. 267-272, 1980.
- ALVAREZ-BUYLLA, E. R.; GARCIA-BARRIOS, R.; LARA-MORENO, C.; MARTÍNEZ-RAMOS, M. Demographic and genetic models in conservation biology. *Annual Review of Ecology and Systematics*, v. 27, p. 387-421, 1996.
- ANDRUS, D. F.; MCGILLARD, L. D. Selection of dairy cattle for overall excellence. *Journal of Dairy Science*, v. 58, p. 1876-1879, 1975.
- APIOLAZA, L. A.; GILMOUR, A. R.; GARRICK, D. J. Variance modelling of longitudinal height data from a *Pinus radiata* progeny test. *Canadian Journal of Forestry Research*, v. 30, p. 645-654, 2000.
- ARAGHI, M.; POLLOTT, G. E.; CURRAN, M. K. Estimating genetic progress in a Suffolk sheep group breeding scheme. *Proceedings of the British Society of Animal Production*, v. 53, p. 174, 1997.
- ARAÚJO, J. A.; SOUSA, R.; LEMOS, L.; BORRALHO, N. M. G. Estimates of genetic parameters and prediction of breeding values for growth in *Eucalyptus globulus* combining clonal and full-sib progeny information. *Silvae Genetica*, v. 45, n. 4, p. 223-226, 1996.
- ARENDONK, J. A. M. van; TIER, B.; KINGHORN, B.P. Use of multiple genetic markers in prediction of breeding values. *Genetics*, v. 137, p. 319-329, 1994.
- ARENDONK, J.A.M. van; ROSMEVLEN, C. van; JANSS, L.L.G.; KNOL, E.F. Estimation of direct and maternal genetic (co)variances for survival within litters of piglets. *Livestock Production Science*, v. 46, p. 163-171, 1996.

- ARNOLD, J. W.; BERTRAND, J. K.; BENYSHEK, K. Animal model for genetic evaluation of multibreed data. **Journal of Animal Science**, v. 70, p. 3322-3332, 1992.
- ASLAMINEJAD, A. A.; RODEN, J. A. Estimation of direct and maternal genetic parameters for 12 week weight for Wens Mountain Sheep. **Proceedings of the British Society of Animal Production**, v. 53, p. 177, 1997.
- ASSIS, T. F. Production and use of *Eucalyptus* hybrids for industrial purposes. In: DUNGEY, H.S.; DIETERS, M.J.; NIKLES, D.G. comp. **Hybrid Breeding and Genetics of Forest trees**: Proceedings of QFRI/CRC-SPF Symposium, 9-14 April 2000, Noosa. Brisbane: Department of Primary Industries, 2000. p.63-74.
- AZEVEDO, D. M. P.; CRISÓSTOMO, J. R.; ALMEIDA, F. C. G.; ROSSETTI, A. G. Estimates of genetic correlations and correlated responses to selection in cashew. **Genetics and Molecular Biology**, v. 21, n. 3, p. 399-402, 1998.
- BAKER, L. H.; CURNOW, R. N. Choice of population size and use of variation between replicate populations in plant breeding selection programs. **Crop Science**, v. 9, p. 555-560, 1969.
- BAKER, R. J. **Selection indices in plant breeding**. Boca Raton: CRC Press, 1986. 218 p.
- BALMELLI, G. Genetic parameter estimates for growth traits in *Eucalyptus grandis*: some implications for breeding and commercial seed production. In: IUFRO CONFERENCE ON FOREST GENETICS FOR THE NEXT MILLENIUM, 2000, Durban. **Proceedings...** Durban: IUFRO / ICFR, 2000. p. 65-68.
- BARADAT, P. **Génétique quantitative**: modèles statistiques et génétiques de base. Paris: INRA, Amélioration de Arboles Forestiers, 1982. 205p.
- BARADAT, P. LABBÉ, T.; BOUVET, J. M. Conception d'index pour la sélection réciproque récurrent: aspects génétiques, statistiques et informatiques. In: CIRAD (Montpellier) **Traitements statistiques des essais de sélection**: stratégies d'amélioration des plantes pérennes. Actes du Séminaire de Biométrie et Génétique Quantitative. Montpellier, 1995. p. 101-150.
- BARBIN, D. **Componentes de variância**. Piracicaba: ESALQ, 1993, 112 p.
- BARBOSA, M. H. P. Perspectivas para o melhoramento da cana-de-açúcar In. IV SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 2000, Lavras. Genética e Melhoramento de Espécies de Propagação Vegetativa. **Anais**. Lavras: UFLA, 2000. p. 1-17.
- BARCELOS, E.; AMBLARD, P. **Oil Palm Breeding Program at EMBRAPA/BRAZIL**. Manaus: EMBRAPA-CPAA, 1992. 20 p.
- BARCELOS, E. ; NUNES, C. D. M.; CUNHA, R. N. V. Melhoramento genético e produção de sementes comerciais de dendêzeiro. In: VIEGA, I. J. M.; MULLER, A. A. **A cultura do dendêzeiro na Amazônia Brasileira**. Belém: Embrapa, 2000. p. 145-174.

BARDORFF-NIELSEN, O. Plausibility inference. *Journal of the Statistical Society, Series B*, v. 38, p. 103-131, 1976.

BARIL, C.; BERGONZINI, J.C. Appréciation de la valeur génétique. In: CIRAD (Montpellier) *Traitements statistiques des essais de sélection*: stratégies d'amélioration des plantes pérennes. Actes du Séminaire de Biométrie et Génétique Quantitative. Montpellier, 1995a. p. 25-38.

BARIL, C.; BERGONZINI, J.C. Le modèle quantitatif. In: CIRAD (Montpellier) *Traitements statistiques des essais de sélection*: stratégies d'amélioration des plantes pérennes. Actes du Séminaire de Biométrie et Génétique Quantitative. Montpellier, 1995b. p. 47-55.

BARKER, J. E.; LIBBY, W. J. The use of selfing in selection of forest trees. *Journal of Genetics*, v. 61, p. 152-167, 1974.

BARKER, J. S. F. Population structure. In: HILL, W. G.; MACKAY, T. F. C. *Evolution and animal breeding*. Wallingford: CAB International, 1989. p. 75-82.

BARNES, R. D. Multiple population tree breeding in Zimbabwe. In: IUFRO CONFERENCE ON BREEDING THEORY, PROGENY TESTS AND SEED ORCHARDS, 1986, Williamsburg. *Proceedings...* Williamsburg: Iufro, 1986. p. 285-297.

BARNES, R. D.; BURLEY, J.; GIBSON, G. L.; LEON, J. P. G. Genotype - environment interactions in tropical pines and their effects on the structure of breeding populations. *Silvae Genetica*, v. 33, p. 186-198, 1984.

BARNES, R. D.; MULLIN, L. J. The multiple population breeding strategy in Zimbabwe: five year results. In: IUFRO CONFERENCE ON BREEDING TROPICAL TREES: Population Structure and Genetic Improvement Strategies in Clonal and Seedling Forestry, 1989, Pattaya. *Proceedings....* Oxford: Oxford Forestry Institute, 1989. p. 148-158.

BARNES, R. D.; SIMONS, A. J.; MACQUEEN, D. J. Domestication of hardwood tree species for non-industrial use. In: LAMBETH, C.; DVORAK, W. *Breeding tropical trees*. IUFRO CONFERENCE, 1992, Cali. Proceedings. Raleigh: North Carolina State University, 1992. p. 372-384.

BARROS, L. M.; CRISÓSTOMO, J. R. Melhoramento genético do cajueiro. In: ARAUJO, J. P. P.; SILVA, V. V. (Ed.). *Cajucultura*: modernas técnicas de produção. Fortaleza: Embrapa-CNPPAT, 1995, p. 73-93.

BARROS, L. M.; PAIVA, J. R.; CRISOSTOMO, J. R.; CAVALCANTI, J.J.V. Hibridação em cajueiro. In: BOREM, A. (Ed.). *Hibridação artificial em plantas*. Viçosa: Ed. da UFV, 1999. p.191-220.

BARWICK, S. Introducing economics to modern animal breeding. In: HAMMOND, K.; GRASER, H-U.; Mc DONALD, C. A. *Animal breeding*: the modern approach. Sydney: Post Graduate Foundation in Veterinary Science, 1992. p. 121-140.

- BAUDOUIN, L. Sélection récurrente réciproque chez le palmier à huile. In: CIRAD (Montpellier) **Traitements statistiques des essais de sélection**: stratégies d'amélioration des plantes pérennes. Actes du séminaire de biométrie et génétique quantitative. Montpellier: Cirad-CP, 1995. p. 217-226.
- BAUDOUIN, L.; BARIL, C.; CLEMENTDEMANGE, A.; LEROY, T.; PAULIN, D. Recurrent selection of tropical tree crops. **Euphytica**, v. 96, n. 1, p. 101-114, 1997.
- BAUDOUIN, L.; DURANDGASSELIN, T. Genetic transmission of characters linked to oil yields in oil palm by cloning results for young palms. **Oleagineux**, v. 46, n. 8-9, p. 313-320, 1991.
- BAYES, T. An essay towards solving a problem in the doctrine of chances. **Philos. Trans. R. Soc.**, London, v. 53, p. 370-418, 1763.
- BECKER, W. **A manual of quantitative genetics**. 4.ed. Pullman: Academic Enterprises, 1984. 196 p.
- BELINGHERI, L. D.; PRAT KRICUN, S. D. Programa de mejoramiento genetico de la yerba mate en el INTA. In: CONGRESSO SUL-AMERICANO DA ERVA MATE, 2.; REUNIAO TECNICA DA ERVA MATE, 2.; 1997, Curitiba. **Anais...** Colombo: Embrapa, 1997. p. 267-278.
- BELONSKY, G. M.; KENNEDY, B. W. Selection on individual phenotype and best linear unbiased predictor of breeding value in a closed swine herd. **Journal of Animal Science**, v. 66, p. 1124-1131, 1988.
- BENNETT, C. A.; FRANKLIN, N. L. **Statistical analysis in chemistry and the chemical industry**. New York: John Wiley & Sons, 1963. 724p.
- BERK, K. Computing for incomplete repeated measure. **Biometrics**, v. 43, n. 1, p. 385-398, 1987.
- BERNARDO, R. Best linear unbiased prediction of maize single cross performance. **Crop Science**, v. 36, p. 50-56, 1996.
- BERNARDO, R. Best linear unbiased predictor analysis. In: COORS, J.G.; PANDEY, S. (Ed.). **The genetics and exploitation of heterosis in crops**. Madison: ASA / CSSA / SSSA, 1999. p. 269-276.
- BERTOLUCCI, F. L. G.; PENCHEL, R. Clonagem do eucalipto: efeitos sobre a produtividade e qualidade da madeira. **Ciência Hoje**, v. 16, n. 91, p.16-21, 1993.
- BHATT, G. M. Comparison of various methods of selecting parents for hybridization in common bread wheat (*Triticum aestivum* L.). **Australian Journal of Agricultural Research**, v. 24, n. 4, p. 457-64, 1973.
- BI, G. C.; YUAN, L. J.; WILLIAMS, E. R. Performance of *Acacia mearnsii* provenances / progeny in Southern China. In: TURNBULL, J. W. **Advances in tropical acacia research**. Canberra: Australian Centre for International Agricultural Research, 1991. p. 215-218. (ACIAR. Proceedings, 35).

- BIBBY, J.; TOUTENBURG, H. **Prediction and improved estimation in linear models.** Chichester: John Wiley and Sons, 1977. 188 p.
- BICHARD, M. Dissemination of genetic improvement through a livestock industry. **Animal Production**, v. 13, p. 401-411, 1971.
- BINET, F. E. On the construction of an index for indirect selection. **Biometrics**, v. 21, p. 291-299, 1965.
- BLISS, C. I. The calculation of the dosage mortality curve. **Animals of Applied Biology**, v. 22, p. 134-167, 1935.
- BLISS, C. I. The method of probits. **Science**, v. 79, p. 38, 1934.
- BOER, I. J. M.; HOESCHELE, I. Genetic evaluation methods for populations with dominance and inbreeding. **Theoretical and Applied Genetics**, v. 86, p. 245-258, 1993.
- BOHREN, B. B.; HILL, W. G.; ROBERTSON, A. Some observations on asymmetrical correlated responses to selection. **Genetical Research**, v. 7, p. 44-57, 1966.
- 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.
- BONNOT, F. Dispositifs expérimentaux et taille des parcelles dans les essais de sélection sur plantes pérennes. In: CIRAD (Montpellier). **Traitements statistiques des essais de sélection:** stratégies d'amélioration des plantes pérennes. Montpellier, 1995. p. 161-172. Actes du Séminaire de Biométrie et Génétique Quantitative.
- BORÉM, A. **Melhoramento de plantas.** 2.ed. Viçosa: UFV, 1997. 453 p.
- BORGES, R. C. G. **Estimativas de herdabilidade e correlações entre caracteres de crescimento em Eucalyptus grandis.** Viçosa: UFV, 1980. 42 p. Tese Mestrado.
- BORGES, R. C. G.; BRUNE, A. Estudo da herdabilidade quanto à resistência a *Diaporthe cubensis* em *Eucalyptus grandis*. **Revista Árvore**, v. 5, n. 1, p. 115-120, 1981.
- BORRALHO, N. M. G. Seed orchards or cuttings: which is the best? In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCALYPT. **Proceedings...** Colombo: EMBRAPA, 1997, v. 1, p. 330-336.
- BORRALHO, N. M. G.; JARVIS, S. F.; POTTS, B. Multivariate restricted maximum likelihood estimation of genetic parameters for diameter at 3 to 5 years in the *Eucalyptus globulus* base population in Australia. In: POTTS, B. et al. **Eucalypt plantations:** improving fiber yield and quality. Hobart: CRC / IUFRO, 1995. p. 230-232.

- BORRALHO, N. M. G.; KANOWSKI, P. J.; COTERRIL, P. P. Genetic control of growth of *Eucalyptus globulus* in Portugal. I Genetic and phenotypic parameters. *Genetica*, Frankfurt, v. 41, n. 1, p. 33-45, 1992.
- BOURDEIX, R; MEUNIER, J.; N'CHO, Y. P. Une stratégie de sélection du cocotier *Cocos nucifera*. II. Amélioration des hybrides Grand x Grand. *Oléagineux*, v. 46, n. 7, p. 267-282, 1991a.
- BOURDEIX, R; MEUNIER, J.; N'CHO, Y. P. Une stratégie de sélection du cocotier *Cocos nucifera*. II. Amélioration des hybrides Nain x Grand. *Oléagineux*, v. 46, n. 10, p. 361-374, 1991b.
- BOUVET, J. M. Selection precoce de l'eucalyptus au Congo. *Bois et Forêts des Tropiques*, v. 246, n. 4, p. 23-40, 1995.
- BOUVET, J. M.; COUTEAU, N.; VIGNERON, P. Premiers éléments de l'analyse des plans factoriels du schéma de sélection récurrent réciproque de l'eucalyptus au Congo. In: PRODUCTION DE VARIÉTÉS GÉNÉTIQUEMENT AMÉLIORÉES D'ESPÈCES FORESTIÈRES À CROISSANCE RAPIDE, 1992. Bordeaux. *Proceedings...* Nangis: AFOCEL, 1992, p. 317-326.
- BOUVET, J. M.; VIGNERON, P. Age trends in variances and heritabilities in Eucalyptus factorial mating designs. *Silvae Genetica*, Frankfurt, v. 44, n. 4, p. 206-216, 1995.
- BOX, G. E. P.; JENKINS, G. M. *Time series analysis*: forecasting and control. San Francisco: Holden-Day, 1970.
- BOX, G. E. P.; TIAO, G. C. *Bayesian inference in statistical analysis*. Reading: Addison-Wesley Publ. Co., 1973. 588 p.
- BRESLOW, N. E.; CLAYTON, D. G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, v. 88, p. 9-25, 1993.
- BREWBAKER, J. L. *Leucaena*: a multipurpose tree genus for tropical agroforestry. In: STEPPLER, H. A.; NAIR, P. K. (Ed.). *Agroforestry: a decade of development*. Nairobi: ICRAF, 1987. p. 289-323.
- BREWBAKER, J. L. Systematics, self-incompatibility, breeding systems and genetic improvement of Leucaena species. In: INTERNATIONAL DEVELOPMENT RESEARCH CENTRE, Ottawa, CA. *Leucaena research in the Asian- Pacific region*: proceedings of a Workshop. 1983. p. 17-22.
- BREWBAKER, J. L. The genetic vulnerability of single variety plantations of Leucaena. *Leucaena Research Reports*, v. 6, p. 81, 1985.
- BRIDGEWATER, F. E. Mating designs. In: FINS, L.; FRIEDMAN, S.T.; BROTSCHOL, J.V. (Ed.). *Handbook of quantitative forest genetics*. Kluwer, 1992, p. 69-95.

BRIQUET JUNIOR, R. ***Melhoramento genético animal***. São Paulo: Melhoramentos / USP, 1967. 269 p.

BROWN, A. H. D.; MATHESON, A. C.; ELDRIDGE, K. G. Estimates of the mating system of *Eucalyptus obliqua* using allozyme polymorphisms. ***Australian Journal of Botany***, v. 23, p. 931-949, 1975.

BRUCKNER, C. H. Melhoramento de fruteiras. In: BORÉM, A. (Ed.). ***Melhoramento de espécies cultivadas***. Viçosa: Ed. da UFV., 1999. p. 679-714.

BUENO FILHO, J. S. S. ***Seleção combinada versus seleção sequencial no melhoramento de populações florestais***. Piracicaba: ESALQ, 1992. 96 p. Tese Mestrado.

BUENO FILHO, J. S. S. ***Uso de modelos mistos na predição de valores genéticos aditivos em testes de progêneres florestais***. Piracicaba: ESALQ, 1997. 118 p. Tese Doutorado.

BUENO FILHO, J. S. S. ; VENCOVSKY, R. Alternativas de análise de ensaios em látice no melhoramento vegetal. ***Pesquisa Agropecuária Brasileira***, v. 35, n.2, p. 259 – 269, 2000.

BUIJTENEN, J.P. van. Efficiency of mating designs for second-generation selection. In: IUFRO CONFERENCE ON PROGENY TESTING, 1972, Macon. ***Proceedings...*** Macon: IUFRO, 1972. p. 1-12.

BUIJTENEN, J.P. van. Mating designs. In: IUFRO CONFERENCE ON ADVANCED GENERATION BREEDING, 1976, Bordeaux. ***Proceedings...*** Bordeaux: IUFRO, 1976. p.11-27.

BUIJTENEN, J.P. van; BURDON, R.D. Expected efficiencies of mating designs for advanced-generation selection. ***Canadian Journal of Forestry Research***, v.20, p.1648-1663, 1990.

BUIJTENEN, J.P. van; LOWE, W.J. The use of breeding groups in advanced generation breeding. In: SOUTHERN FOREST TREE IMPROVEMENT CONFERENCE, 15., Mississippi, 1979. ***Proceedings...*** Mississippi: SFTIC, 1979. p. 7-13.

BULMER, M. G. The effect of selection on genetic variability. ***American Naturalist***, v.105, p. 201-211, 1971.

BULMER, M. G. ***The mathematical theory of quantitative genetics***. Oxford: Charedon Press, 1980. 254 p.

BURDON, R. D. Testing and selection: strategies and tactics for the future. In: IUFRO CONFERENCE ON BREEDING TROPICAL TREES, 1992, Cali. ***Proceedings...*** Raleigh: North Carolina State University, 1992. p. 249-260.

BURDON, R. D.; BUIJTENEN, J. P. van. Expected efficiencies of mating designs for reselection of parents. ***Canadian Journal of Forestry Research***, v. 20, p. 1664-1671, 1990.

- BURDON, R. D.; NAMKOONG, G. Multiple population and sublines. *Genetica*, Frankfurt, v. 32, p. 221-222, 1983.
- BURDON, R. D.; SHELBOURNE, C. J. A. Breeding populations for recurrent selection: conflicts and possible solutions. *New Zealand Journal of Forest Science*, v. 1, p. 174-193, 1971.
- BURDON, R. D.; SHELBOURNE, C. J. A. The use of vegetative propagules for obtaining genetic information. *New Zealand Journal of Forest Science*, v. 4, n. 2, p. 418-425, 1974.
- BURLEY, I. Comments on "Handbook of Quantitative Forest Genetics". *Commonwealth Forestry Review*, v. 72, n. 2, p. 128, 1993.
- BURROWS, P. M. Expected selection differentials for directional selection. *Biometrics*, v. 28, p. 1091-1100, 1972.
- BURROWS, P.M. Inbreeding under selection from unrelated families. *Biometrics*, v. 40, p. 357-366, 1984.
- BUTLER, I.; WILLEKE, H.; PIRCHNER, F. Two-way within family and mass selection for 8-week body weight in different mouse populations. *Genetic Research*, v. 43, p. 191-200, 1984.
- BUZAS, J. S. Fast estimators of the jackknife. *The American Statistician*, v. 51, p. 235-240, 1997.
- CABALLERO, A. Developments in the prediction of effective population size. *Heredity*, v. 73, p. 657-679, 1994.
- CABALLERO, A. Efficiency in prediction of response in selection index experiments. *Journal of Animal Breeding and Genetics*, v. 106, p. 187-194, 1989.
- CAMCORE (Raleigh, North Carolina). *Annual Report*. Raleigh: North Carolina State University, 1997. 23 p.
- CAMERON, N. D. *Selection indices and prediction of genetic merit in animal breeding*. Wallingford: CAB International, 1997. 203 p.
- CAMPOS, H. *Estatística experimental não-paramétrica*. Piracicaba: ESALQ, 1983. 349 p.
- CARABAÑO, M. J.; DÍAZ, C. Predicción del valor genético: BLUP-Modelo Animal y sus aplicaciones. In: BUXADÉ, C. (Ed.). *Zootecnia: bases de producción animal*. Madrid: Mundi-Prensa, 1995. v. 4, p. 123-135.
- CARDELLINO, R. A.; ROVIRA, J. *Mejoramiento genético animal*. Montevideo: Hemisferio Sur, 1987, 253p.
- CARRICK, M.; ENGLAND, D. P. The Merinotech structure and across-flock breeding value prediction. *Proceedings of the Australian Society of Animal Production*, v. 8, p. 223-226, 1990.

CARVALHEIRO, R.; FRIES, L. A.; MUNIZ, C. A. S. D.; QUEIRÓZ, S. A. Estudo de simulação das relações entre a média aritmética, a média harmônica, e o desvio padrão do ganho médio diário de peso do nascimento ao desmame de bovinos. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 38., 2001, Piracicaba. **Anais...** Piracicaba: FEALQ, 2001. CD ROM.

CASTRO, N. H. C. de. **Número de repetições e eficiência da seleção em progêniess de meios-irmãos de Eucalyptus camaldulensis**. Lavras: ESAL, 1992. 96 p. Tese Mestrado.

CAVALCANTI, J. J. V.; PINTO, C. A. B. P.; CRISÓSTOMO, J. R.; FERREIRA, D. F. Análise dialélica para avaliação de híbridos interpoplacionais de cajueiro. **Pesquisa Agropecuária Brasileira**, v. 35, n. 8, p. 1567-1575, 2000.

CHAMBERS, P. G. S.; BORRALHO, N. M. G.; POTTS, B. M. Genetic analysis of survival in *Eucalyptus globulus* ssp. *globulus*. **Silvae Genetica**, Frankfurt, v. 45, n. 2-3, p. 107-112, 1996.

CHEW, V. Multiple comparison procedures: uses, abuses and alternatives. In: WORKSHOP OF THE GENETICS AND BREEDING OF SOUTHERN FOREST TREES, 1986, Gainesville. **Proceedings...** Gainesville: University of Florida, 1986. p.48-58.

CILAS, C. Dispositifs expérimentaux adaptés aux essais de sélection chez le cacaoyer. In: CIRAD (Montpellier). **Traitements statistiques des essais de sélection**: stratégies d'amélioration des plantes pérennes.. Montpellier, 1995a. p.151-160. Actes du Séminaire de Biométrie et Génétique Quantitative.

CILAS, C. Estimation des variances génétiques et des héritabilités pour différents plans de croisements. In: CIRAD (Montpellier). **Traitements statistiques des essais de sélection**: Stratégies d'amélioration des plantes pérennes. Montpellier, 1995b. p. 71-88. Actes du Séminaire de Biométrie et Génétique Quantitative.

CILAS, C.; BOUHARMONT, P.; BOCCARA, M.; ESKES, A. B.; BARADAT, P. Prediction of genetic value for coffee production in Coffea arabica from a half-diallel with lines and hybrids. **Euphytica**, v. 104, p. 49-59, 1998a.

CILAS, C.; LANAUD, C.; PAULIN, D; NYASSÉ, S; N'GORAN, J. A.; KÉBÉ, B. I.; DUCAMP, M.; FLAMENT, M. H.; RISTERUCCI, A. M.; PIERETTI, I.; SOUNIGO, O.; THÉVENIN, J. M.; DESPRÉAUX, D. La résistance à la pourriture des cabosses due à *Phytophthora* spp. Recherche des composantes de la résistance. **Plantations, Recherche, Développement**, v. 5, p. 441-449, 1998b.

CIRAD (Montpellier). **Traitements statistiques des essais de sélection**: stratégies d'amélioration des plantes pérennes. Montpellier, 1995. 385 p. Actes du séminaire de biométrie et génétique quantitative.

CLAYTON, G. A.; MORRIS, J. A.; ROBERTSON, A. An experimental check on quantitative genetical theory. I. Short-term responses to selection. **Journal of Genetics**, v. 55, p. 131-151, 1957.

CLEMENT, C. R.; BOVI, M. L. A. Melhoramento genético da pupunheira: conhecimentos atuais e necessidades. In: SEMINÁRIO DO AGRONEGÓCIO PALMITO DE PUPUNHA NA AMAZÔNIA, 1., 1999, Porto Velho. *Anais...* EMBRAPA-CPAF Rondônia, 1999. p. 57-70. (EMBRAPA-CPAF Rondônia. Documentos, 41).

CLEMENT, C. R.; CHAVEZ, W. B.; GOMES, J. B. M. Considerações sobre a pupunha como produtora de palmito. In: ENCONTRO NACIONAL DE PESQUISADORES EM PALMITO, Curitiba, 1987. Curitiba: Embrapa, 1987. p.225 – 247. (Documentos, 19).

CLEMENT, C. R.; MULLER, C. H.; FLORES, W. B. C. Recursos genéticos de espécies frutíferas nativas da Amazônia brasileira. *Acta Amazônica*, Manaus, v. 12, n. 4, p. 677-695, 1982.

CLEMENT-DEMANGE, A.; ESKES, A.; SOUNIGO, O.; GORAN, J. Amélioration génétique du cacaoyer en Côte d'Ivoire, présentation d'un nouveau schéma de sélection. In: INTERNATIONAL CACOA RESEARCH CONFERENCE, Yamoussoukro, 1993. Yamoussoukro, Côte d'Ivoire, 1993, p. 451-455.

CLÉMENT-DEMANGE, A.; RIVANO, F.; NICOLAS, D.; GNAGNE, H.; LEGNATÉ, H. Stratégies de sélection chez l'hévéa. In: CIRAD (Montpellier). *Traitements statistiques des essais de sélection: stratégies d'amélioration des plantes pérennes*. Montpellier, 1995. p. 227-242. Actes du séminaire de biométrie et génétique quantitative.

COCHARD, B.; NOIRET, J. M.; BAUDOUIN, L.; AMBLARD, P. Second cycle de sélection récurrente réciproque chez le palmier à huile *Elaeis guineensis*. Résultat des tests d'hybrides Deli x La Mé. *Oléagineux*, v. 48, n. 11, p. 441-451, 1993.

COCHRAN, W. G. Improvement by means of selection. SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 2., 1951, Berkeley. *Proceedings...* Berkeley: University of California Press, 1951. p.449-470.

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

COCKERHAM, C. C. Covariances of relatives from self-fertilization. *Crop Science*, v. 23, p. 1177-1180, 1983.

COCKERHAM, C. C. Estimation of genetic variances. In: HANSON, W. D.; ROBINSON, A. F. (Ed.). *Statistical genetics and plant breeding*. Washington: National Academy of Science, 1963. p. 53-93. (NAS-RNC n.982).

COCKERHAM, C. C. Higher order probability functions of identity of alleles by descent. *Genetics*, v. 69, p. 235-246, 1971.

COCKERHAM, C. C.; WEIR, B. S. Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics*, v. 40, p. 157-164, 1984.

COCKERHAM, C.C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. ***Genetics***, v. 39, p. 859-882, 1954.

COMSTOCK, R. E. ***Quantitative genetics with special reference to plants and animal breedings***. Ames: Iowa State University Press, 1996. 421 p.

CORDEIRO, G. M. ***Modelos lineares generalizados***. Campinas: Unicamp (VII SINAPE), 1986. 286 p.

CORNELLIUS, J. Heritabilities and additive genetic coefficients of variation in forest trees. ***Canadian Journal of Forestry Research***, v. 24, p. 371-379, 1994.

COSTA e SILVA, J.; DUTKOWSKI, G. W.; GILMOUR, A. R. Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. ***Silvae Genetica***, v. 31, p. 1887-1893, 2001.

COSTA, J. G.; LEDO, A. S.; OLIVEIRA, M. N. Estimativas de repetibilidade de características de frutos de cupuaçuzeiro no Estado do Acre. ***Revista Brasileira de Fruticultura***, v. 19, n. 3, p. 313-318, 1997.

COSTA, R. B.; RESENDE, M. D. V.; ARAÚJO, A. J.; GONÇALVES, P. S.; HIGA, A. R. Selection and genetic gain in rubber tree (*Hevea*) populations using a mixed mating system. ***Genetics and Molecular Biology***, Ribeirão Preto, v. 23, n. 3, p. 671-679, 2000.

COTTERILL, P. P. JAMES, J. Number of offspring and plot sizes required for progeny testing. ***Silvae Genetica***, Frankfurt, v. 23, n. 6, p. 203-208, 1984.

COTTERILL, P. P. Genetic gains expected from alternative breeding strategies including simple low cost options. ***Silvae Genetica***, Frankfurt, v. 34, n.5-6, p. 212-223, 1986.

COTTERILL, P. P. Nucleus breeding: is it OK for *Pinus* and *Eucalyptus*? In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCLYPT, Salvador, 1997. ***Proceedings...*** Colombo: EMBRAPA, 1997, v.1, p. 277-284.

COTTERILL, P. P. The nucleus-breeding system. In: TWENTIETH SOUTHERN FOREST TREE IMPROVEMENT CONFERENCE, Charleston, 1989. ***Proceedings...*** Charleston: SFTIC, 1989. p. 26-30.

COTTERILL, P. P.; BROLIN, A. Improving Eucalyptus wood, pulp and paper quality by genetic selection. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCLYPTS, 1997, Salvador. [***Proceedings...***]. Co-lombo: EMBRAPA-CNPF, 1997. v.1, p.1-13, 1997.

- COTTERILL, P. P.; DEAN, C. A. ***Successful tree breeding with index selection.*** Melbourne: CSIRO, 1990. 80 p.
- COTTERILL, P. P.; DEAN, C. A.; WYK, G. van. Additive and dominance genetic effects in *Pinus pinaster*, *P. radiata* e *P. elliottii* and some implications for breeding strategies. ***Silvae Genetica***, v. 36, n.5-6,p. 221-232, 1987.
- COTTERILL, P. P.; JACKSON, N. Index selection with restrictions in tree breeding. ***Silvae Genetica***, Frankfurt, v. 30, n.2-3, p. 106-108, 1981.
- COTTERILL, P. P.; JAMES, J. W. Optimising two-stage independent culling selection in tree and animal breeding. ***Theoretical and Applied Genetics***, v. 59, p. 67-72, 1981.
- COTTERILL, P.P.; DEAN, C.A.; CAMERON, J.; BRINDBERGS, M. Nucleus breeding: A new strategy for rapid improvement under clonal forestry. In: IUFRO MEETING ON BREEDING TROPICAL TREES, Pattaya, 1989. ***Proceedings...*** Oxford: Oxford Forestry Institute, 1989. p. 1-15.
- COTTERILL, P.P.; JACKSON, N. On index selection. I. Methods of determining economic weight. ***Silvae Genetica***, Frankfurt , v. 34, n.2-3, p. 56-63, 1985.
- CRESSIE, N. A. C. ***Statistics for spatial data analysis.*** New York: J. Wiley & Sons, 1993. 900 p.
- CRISÓSTOMO, J. R. **Avaliação da estrutura e do potencial genético de uma população de algodoeiro (*C. hirsutum* L) parcialmente autógama.** Piracicaba: ESALQ, 1989. 191 p. Tese Doutorado.
- CRISÓSTOMO, J. R.; BARROS, L. de M.; PAIVA, J. R. de ; CAVALCANTI J. J. V. Melhoramento genético do cajueiro. In: QUEIRÓZ, M. A; GOEDERT, C. O; RAMOS, S. R. R. (Ed.) ***Recursos genéticos e melhoramento de plantas para o nordeste brasileiro.*** Petrolina: CPATSA, 1999. Anais on line. Disponível em www.cpatsa.embrapa.br.
- CRITTENDEN, L. B.; BOHREN, B. B.; ANDERSON, V. L. Genetic variance and covariance of the components of hatchability in New Hampshires. ***Poultry Science***, v. 36, p. 90-103, 1957.
- CROSBIE, S. F.; HINCH, G. N. An intuitive explanation of generalised linear models. ***New Zealand Journal of Agricultural Research***, v. 28, p. 19-29, 1985.
- CROW, J. F.; DENNISTON, C. Inbreeding and variance effective population numbers. ***Evolution***, v. 42, n. 3, p. 482-495. 1988.
- CROW, J.F.; KIMURA, M. ***An introduction to population genetics theory.*** New York: Harper and Row, 1970. 591 p.
- CRUZ, C. D.; REGAZZI, O. J. ***Modelos biométricos aplicados ao melhoramento genético.*** Viçosa: Universidade Federal de Viçosa, Imprensa Universitária, 1994. 390 p.

- CRUZ, C. D. **Programa GENES**: aplicativo computacional em genética e estatística. Viçosa: Universidade Federal de Viçosa, Editora UFV, 1997. 442p.
- CULBERTSON, M. S.; MABRY, J. W.; MISZTAL, I.; GENGLER, N.; BERTRAND, J. K. VARONA, L. Estimation of dominance variance in purebred Yorkshire swine. *Journal of Animal Science*, v. 76, p. 448-450, 1998.
- CULLIS, B. R.; GOGELL, B; VERBYLA, A. Spatial analysis of multi-environment early generation variety trials. *Biometrics*, v. 54, p. 1-18, 1998.
- CULLIS, B. R.; GLEESON, A. C. Spatial analysis of field experiments-an extension at two dimensions. *Biometrics*, v. 47, p. 1449-1460, 1991.
- CULLIS, B. R.; WELHAM, S. J.; THOMPSON, R.; GILMOUR, A. R. **New developments in REML and its implementation in Genstat 5 Release 4.1**: Workshop Notes. Melbourne: NSW Agriculture, 1999. p.57.
- CUNNINGHAM, E. P. Multi-stage index selection. *Theoretical and Applied Genetics*, v. 46, p. 55-61, 1975.
- CUNNINGHAM, E. P. O'BYRNE, T.; MESCAL, A. A. Genetic relationship between dairy and beef traits in Freisian cattle. *Irish Journal of Agriculture Research*, v. 16, p. 243-249, 1977.
- CUNNINGHAM, E. P.; MOEN, R. A.; GJEDREM, T. Restriction of selection indexes. *Biometrics*, v. 26, p. 67-74, 1970.
- CURNOW, R. N.; BAKER, L. H. The effect of repeated cycles of selection and regeneration in populations of finite size. *Genetical Research*, Cambridge, v. 11, p. 105-112, 1968.
- CURRIE, A. J.; GANESHANANDAM, S.; NOITON, D. A; GARRICK, D.; SHELBORNE, C. J. A.; ORAGUZIE, N. Quantitative evaluation of apple fruit shape by principal component analysis of Fourier descriptors. *Euphytica*, v. 111, n. 3, p. 219-227, 2000.
- DAMODARAN, V. K. Hybrid vigour in cashew. *Agricultural Research Journal of Kerala*, v. 13, n. 2, p.195-196, 1975.
- DANEL, O. **H₄REML, a FORTRAN computer programme for variance component estimation and BLUP evaluation of data with a stand and half-sib family structure**. Uppsala: Institute for Forest Improvement, 1987.
- DANNEL, O. Advanced models of predicting selection gain. In: WORLD FOREST CONGRESS, TAMPERE, 1995. **Proceedings...** Tampere: IUFRO, 1995, p. 152.
- DAVIS, J. C. **Statistics and data analysis in geology**. New York: J. Wiley, 1986. 646 p.
- DEAN, C. A.; COTTERILL, P. P.; EISEMANN, R. L. Genetic parameters and gains expected from selection in *Pinus caribaea* var. *hondurensis* in Northern Queensland, Austrália. *Silvae Genetica*, v. 35, n. 5-6, p. 229-236, 1986.

- DEGEN, B.; SCHOLZ, F. Spatial genetic differentiation among populations of european beech in western Germany as identified by geostatistical analysis. *Forest Genetics*, v. 5, n. 3, p. 191-199, 1998.
- DEL-BOSQUE-GONZALES, A. S.; KINGHORN, B. P. Consequences of genotype x environment interaction in group breeding schemes. In: MCGUIRK, B.J. (Ed.). *Merino Improvement Programs in Australia*. Melbourne: Australian Wool Corporation, 1987. p. 437-442.
- DEMÉTRIO, C. G. B. *Modelos lineares generalizados na experimentação agronômica*. Porto Alegre: Universidade Federal do Rio Grande do Sul, 1993. 125 p.
- DEMPFLE, L. Relation entre BLUP (Best linear unbiased prediction) et estimateurs bayesiens. *Annales de Génétique et Sélection Animale*, v. 9, p. 27-32, 1977.
- DEMPFLE, L. A note on increasing the limit of selection through selection within families. *Genetical Research*, v. 24, p. 127-135, 1975.
- DEMPFLE, L. Estimation of breeding values. In: HILL, W. G.; MACKAY, T. F. C. (Ed.). *Evolution and animal breeding*. Wallingford: CAB International, 1989. p. 181-188.
- DEMPFLE, L. Statistical aspects of design of animal breeding programmes: a comparison among various selection strategies. In: GIANOLA, D. HAMMOND, K. (Ed.). *Advances in statistical methods for genetic improvement of livestock*. Berlin: Springer-Verlag, 1990. p. 109-128.
- 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.
- DEMPSTER, E. R.; LERNER, I. M. Heritability of threshold characters. *Genetics*, v. 35, p. 212-236, 1950.
- DENARDI, F.; CAMILO, A. P. Epagri 406-Baronesa: nova cultivar de macieira de maturação tardia para o Sul do Brasil. *Revista Brasileira de Fruticultura*, v. 1 9, n. 2, p. 185-189, 1997.
- DENARDI, F.; CAMILO, A. P. Estratégia do melhoramento genético da macieira na EPAGRI, Santa Catarina. In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 2., 1998, Lavras. *Anais...* Lavras: UFLA, 1998. p.123-133.
- DIAS, L. A. S. dos. *Melhoramento genético do cacauzeiro*. Viçosa: FUNAPE, 2001. 578 p.
- DIAS, L. A. S. dos. Novos rumos no melhoramento genético do cacauzeiro. In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 2., 1998, Lavras. *Anais...* Lavras: Editora UFLA, 1998a. p. 9-27.

DIAS, L. A. S. dos. Variâncias de frequências alélicas. In: ALFENAS, A. C. (Ed.) **Eletroforese de isoenzimas e proteínas afins**: fundamentos e aplicações em plantas e microorganismos. Viçosa : Ed. UFV, 1998b. p. 381-404.

DIAS, L. A. S. dos; RESENDE, M. D. V. de. Estratégias e métodos de seleção. In: DIAS, L. A. S. dos. (Org.). **Melhoramento genético do cacaueiro**. Viçosa: FUNAPE, 2001b. p. 217-287.

DIAS, L. A. S. dos; RESENDE, M. D. V. Experimentação no melhoramento. In: DIAS, L. A. S. dos. (Org.). **Melhoramento genético do cacaueiro**. Viçosa: FUNAPE, 2001a, p. 439-492.

DIAS, L. A. S. dos; RESENDE, M. D. V. de. Estimation of genetic parameters and prediction of breeding values by mixed model in the cacao improvement. In: INTERNATIONAL COCOA RESEARCH CONFERENCE, 13., 2000, Kota Kinabalu. **Proceedings...** Kuala Lumpur: Cocoa Producer's Alliance, 2001c. p. 189-194.

DIAS, L. A. S. dos; SOUZA, C. A. S. Aplicação do coeficiente de repetibilidade na seleção de cacaueiros em plantação comercial. **Revista Brasileira de Genética**, v. 16, n. 3, p. 364, 1993.

DICKERSON, G. E. Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. **Iowa Agriculture Experimental Station**, v. 354, p. 489-524, 1947.

DICKERSON, G. E. Implications of genotype – environmental interactions in animal breeding. **Animal Production**, v. 4, p. 47-63, 1962.

DICKERSON, G. E.; HAZEL, L. N. Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. **Journal of Agricultural Research**, v. 69, n. 12, p. 459-476, 1944.

DICKERSON, G.E. Techniques for research in quantitative animal genetics. In: DICKERSON, G.E. **Techniques and procedures in Animal Production Research**. New York: American Society of Animal Science, 1959. p. 571-605.

DOBSON, A. J. **An introduction to generalized linear models**. Melbourne: Chapman & Hall, , 1990. 174 p.

DORAN, J. C.; CARTER, A. S.; MATHESON, A. C. Variation in root strike of Petford Eucalyptus camaldulensis clones. In: MASS PRODUCTION TECHNOLOGY FOR GENETICALLY IMPROVED FAST GROWING FOREST TREE SPECIES, 1992, Bordeaux. **Proceedings...** Nangis: Association Forêt Cellulose, 1992. p. 407-414.

DUARTE, J. B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293 f. Tese (Doutorado) – ESALQ/USP, Piracicaba.

- DUDA, L. L.; RESENDE, M. D. V. de; CANDIDO, H. C. S. Avaliação genética e seleção em *Eucalyptus grandis* na região de Arapoti-PR. *Revista Árvore*, v. 21, n. 4, p. 537-545, 1997.
- DUDEWICZ, E. J. Introduction to statistics and probability. In: DUDEWICZ, E. J. **Ranking and selection procedures**. New York: Holt / Rinehart & Winston, 1976. p. 315-383.
- DUDLEY, J. W.; LAMBERT, R. J. Ninety generations of selection for oil and protein in maize. *Maydica*, v. 7, p. 1-7, 1992.
- DUNGEY, H. S.; KERR, R. J.; DIETERS, M. J. Comparison of genetic parameters estimated using an infinitesimal and a finite locus model in two subtropical pines and their interspecific hybrid. In: IUFRO CONFERENCE ON FOREST GENETICS FOR THE NEXT MILLENIUM, 2000, Durban: **Proceedings...** Durban: IUFRO / ICFR, 2000. p. 98-102.
- DURATEX. Equipe Técnica. Teste de progêneres de *E. grandis* em Itapetininga. In: 7º CONGRESSO FLORESTAL BRASILEIRO, 7., 1993, Curitiba. *Anais...* Curitiba: SBS, 1993, p. 155-157.
- DUREL, C. E.; LAURENS, F.; FOUILLET, A.; LESPINASSE, Y. Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. *Theoretical and Applied Genetics*, v. 96, n. 8, p. 1077-1085, 1998.
- DVORAK, W. S.; DONAHUE, J. K. Camcore Cooperative Research Review, 1980 -1992. Raleigh: North Carolina State University, Department of Forestry, College of Forest Research, Camcore Cooperative, 1992. p. 93.
- EDWARDS, A. W. F. **Likelihood**. Cambridge: Cambridge University Press, 1972. 235 p.
- EFRON, B. Biased versus unbiased estimation. *Advances in Mathematics*, v. 16, p. 259-277, 1975.
- EFRON, B. **The jackknife, the bootstrap and other resampling plans**. Philadelphia: Society for Industrial and Applied Mathematics, 1982. 385p.
- EFRON, B. Why isn't everyone a bayesian? *American Statistician*, v. 40, p. 11, 1986.
- EFRON, B. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, v. 82, p. 171-200, 1987.
- EFRON, B. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, v.7, p. 1-26, 1979.
- EFRON, B.; MORRIS, C. Stein's paradox in statistics. *Scientific American*, v. 236, n. 5, p. 119-127, 1977.

EFRON, B.; TIBSHIRANI, R. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, v. 1, p. 54-77, 1986.

EISEN, E. J. Genetic models to predict crossbreed performance: a review. *Rivista Brasileira de Genética*, v. 12, n. 3, p. 13-26, 1989.

EISEN, E. J.; LEGATES, J. E.; ROBISON, O. W. Selection for 12-day litter weight in swine. *Genetics*, v. 4, p. 511-532, 1970.

EISENBERG, B. E.; GAUCH, H. G.; ZOBEL, R. W.; KILIAN, W. Spatial analysis of field experiments: fertilizer experiments with wheat (*Triticum aestivum*) and tea (*Camellia sinensis*). In: KANG, M. S.; GAUCH, H. G. (Ed.). *Genotype by environment interaction*. Boca Raton: CRC Press, 1996. p. 373-404.

ELDRIDGE, K. G. Breeding systems, variation and genetic improvement of tropical Eucalyptus. In: BURLEY, Y.; STYLES, B. T. (Ed.). *Tropical trees*: variation, breeding and conservation. Oxford: Oxford Forestry Institute, 1976. p. 101-108.

ELDRIDGE, K. G. Genetic improvement of Eucalyptus. *Silvae Genetica*, Frankfurt, v. 27, p. 205-209, 1978.

ELDRIDGE, K.; DAVIDSON, J.; HARWOOD, C.; WYK, G. van. *Eucalypt domestication and breeding*. Oxford: Oxford Science Publications. 1993, 288 p.

ELLSTRAND, N. C.; ELAM, D. R. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics*, v. 24, p. 217-242, 1993.

ELSTON, D.A. Estimating of denominator degrees of freedom of F-distributions for assessing Wald statistics for fixed-effect factors in unbalanced mixed models. *Biometrics*, v. 54, n. 3, p. 1085-1096, 1998.

ELSTON, R. C. Estimating heritability of a dichotomous trait. *Biometrics*, v. 33, p. 231-233, 1977.

ENGEL, B.; KEEN, A. A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, v. 48, n. 1, p. 1-22, 1994.

ENGEL, B.; KEEN, A. An introduction to generalized linear mixed models. In: XVIII INTERNATIONAL BIOMETRIC CONFERENCE: Invited Papers, 18., 1996, Amsterdam. *Proceedings...* Amsterdam: International Biometric Society, p. 125-135.

ERICSSON, T.; DANELL, O.; ANDERSON, B. Genetic variation of *Pinus contorta* var. *latifolia* breeding material in Sweden. *Canadian Journal of Forestry Research*, v. 24, p. 723-729, 1994.

- ESCOBAR, J. R. Herdabilidade de alguns caracteres da fase juvenil de clones de guaraná (*Paullinia cupana* var. *sorbilis*). In: SIMPÓSIO DO TRÓPICO ÚMIDO, 1., 1984, Belem. *Anais...* Belém: CPATU, 1986. p.285-293. (EMBRAPA-CPATU. Documentos, 36).
- EWENS, W.J. *Mathematical population genetics*. Berlin: Springer-Verlag, 1979. 325 p.
- FAIRFULL, R. W.; MUIR, W. M. Quantitative genetics versus molecular genetics in poultry breeding. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 34., 1997, Juiz de Fora. *Anais...* Juiz de Fora: SBZ, 1997. p.52-78.
- FALCONER, D. S. *Introduction to quantitative genetics*. 3. ed. Harlow: Longman, 1989. 438 p.
- FALCONER, D. S. The problem of environment and selection. *American Naturalist*, v. 86, p. 293-298, 1952.
- FALCONER, D. S. Early selection experiments. *Annual Review of Genetics*, v. 26, p. 1-14, 1992.
- 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.
- FARIAS NETO, J. T; RESENDE, M. D. V. de. Aplicação da metodologia de modelos mistos (REML/BLUP) na estimativa de componentes de variância e predição de valores genéticos em pupunheira (*Bactris gasipaes L.*). *Revista Brasileira de Fruticultura*, v. 23, n.2, p. 320-324, 2001.
- FASOULAS, A. *Principles and methods of plant breeding and field experimentation*. Thessaloniki: Aristotelian University, 1981. 147 p.
- FEDERER, W. T. Augmented designs. *Biometrics*, v. 14, p. 134, 1958.
- FEDERER, W. T. *Experimental designs: theory and application*. New Delhi: Oxford Publ., 1955. 544 p.
- FEDERER, W. T. Recovery of interblock, intergradient and intervarietal information in incomplete block and lattice rectangle designed experiments. *Biometrics*, v. 54, p. 471-481, 1998.
- FELLNER, W. H. Robust estimation of variance components. *Technometrics*, v. 28, p. 51-60, 1986.
- FELLNER, W.H. Sparce matrix and the estimation of variance components by likelihood methods. *Communications in statistics: Theory and Methods*, B, v. 16, p. 439-463, 1987.

FELSENSTEIN, J. Alan Robertson's contribution to population genetics. In: HILL, W.G.; MACKAY, T.F.C. (Ed.). *Evolution and animal breeding*. Edinburgh, 1989, p. 3-12.

FERNANDES, J. S. C. *Progressos esperados em linhagens de milho (Zea mays L.) via seleção recorrente intra e interpopulacional*. Piracicaba: ESALQ, 1990. 235p. Tese Doutorado.

FERNANDO, R. L.; GIANOLA, D. Optimal properties of the conditional mean as a selection criterion. *Journal of Animal Science*, v. 59, p. 177, 1984.

FERNANDO, R.L.; GROSSMAN, M. Marker-assisted selection using best linear unbiased prediction. *Genetics Selection Evolution*, v. 21, p. 467-477, 1989.

FERRÃO, R. G.; FONSECA, A. F. A.; FERRÃO, M. A. G. Programas de melhoramento genético de café robusta no Brasil. In: NUMBERG, P.L. et al. (Ed.). In: SIMPOSIO DE ATUALIZACAO EM GENETICA E MELHORAMENTO DE PLANTAS, 3., 1999, Lavras. *Anais...* Lavras: UFLA, 1999. p. 50-65.

FERREIRA, M. Melhoramento e a silvicultura intensiva clonal. *IPEF*, Piracicaba, n. 45, p. 22-30, 1992.

FERREIRA, R. P.; PEREIRA, A. V. Melhoramento de forrageiras. In: BORÉM, A. (Ed.). *Melhoramento de espécies cultivadas*. Viçosa: Ed. da UFV, 1999. p. 649-677.

FERWERDA, F. P.; WIT, F. *Outlines of perennial crop breeding in the tropics*. Wageningen: Landbouroh, 1969. 465 p. (Misc. Paper, 4).

FINNEY, D. J. The consequences of selection for a variate subject to errors of measurements. *Rev. Inst. Intern. Statistique*, v. 24, p. 1-10, 1956.

FIRTH, D. Generalized linear models. In: HINKLEY, D.V., REID, N., SNELL, E.J. *Statistical Theory and Modelling*. London: Chapman & Hall, 1991. p.55-82.

FISHER, R. A. *Statistical methods for research workers*. 1. ed. London: Oliver and Boyd, 1925. 314 p.

FISHER, R. A. *Statistical methods for research workers*. 10. ed. New York: Hafner, 1948. 345 p.

FISHER, R. A. The correlation between relatives on the supposition of mendelian inheritance. *Transaction Royal Society of Edinburgh*, v. 32, p. 399-433, 1918.

FISHER, R. A. On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond. Ser. A*, n. 222, p. 309-368, 1922.

FONSECA, C. E. L.; ESCOBAR, J. R.; BUENO, D. M. Variabilidade de alguns caracteres físicos e químicos do fruto do cupuaçuzeiro. *Pesquisa Agropecuária Brasileira*, v. 25, n. 7, p. 1079-1084, 1990.

- FOSTER, G. S.; SHAW, D. V. Using clonal replicates to explore genetic variation in a perennial plant species. *Theoretical and Applied Genetics*, n. 76, p. 788-794, 1988.
- FOULLEY, J. L.; GIANOLA, D.; IM, S. Genetic evaluation for discrete polygenic traits in animal breeding. In: GIANOLA, D.; HAMMOND, K. (Ed.). *Advances in statistical methods for genetic improvement of livestock*. Berlin: Springer-Verlag, 1990, p. 361-409.
- FOULLEY, J. L.; GIANOLA, D.; IM, S. Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theoretical and Applied Genetics*, v. 73, p. 870, 1987.
- FRANKHAM, R. Conservation genetics. *Annual Review of Genetics*, v. 29, p. 305-327, 1995a.
- FRANKHAM, R. Effective population sizes/adult population size ratios in wildlife: a review. *Genetical Research*, v.66, p.95-107, 1995c.
- FRANKHAM, R. Inbreeding and extinction: a threshold effect. *Conservation Biology*, v. 9, p. 792-799, 1995b.
- FREIRE-MAIA, N. *Genética de populações humanas*. São Paulo: HUCITEC / EDUSP, 1974. 280 p.
- FREITAS, M. L. M.; MORAES, M. L. T.; SANTOS, P. E. T.; VALÉRIO FILHO, W. V. Variação genética em progêneres de Eucalyptus urophylla na região de Selvíria-MS. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCALYPTS, 1997, Salvador. [Proceedings...]. Colombo: EMBRAPA, 1997, v.1, p.403-407.
- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, v. 32, p. 675-701, 1937.
- FRIES, L. A.; SCHENKEL, F. S. Estimation and prediction under a selection model. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, Rio de Janeiro, 1993. *Anais dos Simpósios*. Viçosa: Sociedade Brasileira de Zootecnia, 1993. p.1-20.
- FRIES, L.A. Cruzamentos em gado de corte. In: SIMPÓSIO SOBRE PECUÁRIA DE CORTE, 4., 1996, Piracicaba. *Produção do novilho precoce*: anais. Piracicaba: FEALQ, 1997. p. 109-128, 1997.
- FU, Y.; CLARKE, G. P. Y.; NAMKOONG, G.; YANCHUK, A. D. Incomplete block designs for genetic testing: statistical efficiencies of estimating family means. *Canadian Journal of Forest Research*, v. 28, n. 7, p. 977-986, 1998.
- FUJISHIMA, T.; FREDEEN, H.T. General formulae for estimating heritability in a population with related parents. *Canadian Journal of Genetic and Cytology*, v. 14, p. 549-557, 1972.

- FUKUDA, W. M. G. Melhoramento da mandioca. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: Ed. da UFV, 1999. p. 409-428.
- FUTUYAMA, D. J. **Biologia evolutiva**. 2.ed. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 646 p.
- GALE, J. S.; LAWRENCE, M. J. The decay of variability. In: HOLDEN, J. H. W.; WILLIAMS, J. T. **Crop genetic resources**: conservation and evaluation. London: Allen and Unwin, 1984.
- GALLAIS, A. Amélioration des populations, méthodes de sélection et création des variétés. III. Bases théoriques pour l'étude de la sélection récurrente réciproque. **Annales des Améliorations des Plantes**, v. 28, p. 637-660, 1978.
- GALLAIS, A. Sélection pour plusieurs caractères: Synthèse critique et généralisation. **Annales des Améliorations des Plantes**, v. 23, p. 183-208, 1973.
- GALLAIS, A. **Théorie de la sélection en amélioration des plantes**. Paris: Ed. Masson, 1989. 588p.
- GALLO, J.; KHURI, A. I. Exact tests for the random and fixed effects in an unbalanced mixed-two-way cross-classification model. **Biometrics**, v. 46, n. 3, p. 1087-1095, 1990.
- GALTON, F. **Natural inheritance**. London: MacMillan, 1889. 118 p.
- GAMERMAN, D. **Markov chain Monte Carlo**: stochastic simulation for bayesian inference. CRC Press, 1997. 230 p.
- GAMERMAN, D. **Simulação estocástica via cadeias de Markov**. Caxambu: Associação Brasileira de Estatística, 1996. 196 p.
- GAMERMAN, D.; MIGON, H. S. **Inferência estatística**: uma abordagem integrada. Rio de Janeiro: UFRJ / Instituto de Matemática, 1993. 207 p.
- GARCIA-CORTES, L. A.; SORENSEN, D. On a multivariate implementation of the Gibbs samples. **Genetique, Selection, Evolution**, v. 28, p. 121-126, 1996.
- GARDNER, C.O.; EBERHART, S.A. Analysis and interpretation of the variety cross diallel and related populations. **Biometrics**, v. 22, p. 439-452, 1966.
- GARRICK, D. J.; VLECK, L. D. van. Aspects of selection for performance in several environments with heterogeneous variances. **Journal of Animal Science**, v. 65, p. 409-421, 1987.
- GEARY, D. N. Modelling the covariance structure of repeated measurements. **Biometrics**, v. 45, n. 4, p. 1183-1195, 1989.
- GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, p. 398-409, 1990.

- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the bayesian restoration of imagens. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 6, p. 721-741, 1984.
- GIANOLA, D. Inferences about best, worst and ordered genetic values in a small population with unknown variance. In: CONGRESSO NACIONAL DE GENETICA. Caxambu, 1995. (Palestra).
- GIANOLA, D. Statistics in animal breeding. *Journal of the American Statistical Association*, v. 95, n. 449, p. 296-299, 2000.
- GIANOLA, D. Theory and analysis of threshold characters. *Journal of Animal Science*, v. 54, p. 1079-1096, 1982.
- GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. *Journal of Animal Science*, v. 63, p. 217-244, 1986.
- GIANOLA, D.; FERNANDO, R. L.; IM, S.; FOULLEY, J. L. Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems. *Genome*, Ottawa, v. 31, p. 768-777, 1989.
- GIANOLA, D.; FOULLEY, J. L. Sire evaluation for ordered categorical data with a threshold model. *Genetics, Selection, Evolution*, v. 15, p. 201-224, 1983.
- GIANOLA, D.; FOULLEY, J. L. Variance estimation from integrated likelihood (VEIL). *Genetics, Selection, Evolution*, v. 22, p. 403-417, 1990.
- GIANOLA, D.; GOFFINET, B. Sire evaluation with best linear unbiased predictors. *Biometrics*, v. 38, p. 1085-1088, 1982.
- GIANOLA, D.; HAMMOND, K. *Advances in statistical methods for genetic improvement of livestock*. Berlin: Springer-Verlag, 1990. 534 p.
- GIANOLA, D.; IM, S.; MACEDO, F. W. A framework for prediction of breeding value. In: GIANOLA, D.; HAMMOND, K. (Ed.). *Advances in statistical methods for genetic improvement of Livestock*. Berlin: Springer Verlag, 1990. p. 210-238.
- GILL, J. L. *Design and analysis of experiments in the animal and medical sciences*. Ames: Iowa State University Press, 1988. 301 p.
- GILL, J. L. Repeated measurements: sensitive tests for experiments with few animals. *Journal of Animal Science*, v. 63, p. 943-954, 1986.
- GILL, J. L.; HAFS, H. D. Analysis of repeated measurements of animals. *Journal of Animal Science*, v. 33, n. 2, p. 331-336, 1971.
- GILMOUR, A. R. Post blocking gone too far! Recovery of information and spatial analysis in field experiments. *Biometrics*, v. 56, p. 944-946, 2000.
- GILMOUR, A. R.; ANDERSON, R. D.; RAE, A. L. The analysis of binomial data by a generalized linear mixed model. *Biometrika*, London, v. 72, p. 593-599, 1985.

GILMOUR, A. R.; CULLIS, B. R.; WELHAM, S. J.; THOMPSON, R. *ASREML Reference manual*. Orange: NSW Agriculture, 2000. 218 p.

GILMOUR, A. R.; CULLIS, B. R; FRENSHAM, A. B.; THOMPSON, R. (Co)variance structures for linear models in the analysis of plant improvement data. In: COMPSTAT98 COMPUTATIONAL STATISTICS. *Proceedings...* Heidelberg: Physica - Verlag, 1998. p. 53-64.

GILMOUR, A. R.; THOMPSON, R. Modelling variance parameters in ASREML for repeated measures. In: WORLD CONGRESS ON GENETIC APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. *Proceedings...* Armidale: AGBU / University of New England, 1998. v. 27, p. 453-454.

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.

GILMOUR, A.R.; CULLIS, B. R.; VERBYLA, A. P. Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological and Environmental Statistics*, v. 2, p. 269-293, 1997.

GLEESON, A. C.; CULLIS, B. R. Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics*, v. 43, p. 277-288, 1987.

GODDARD, M. E. A mixed model for analysis of data on multiple genetic markers. *Theoretical and Applied Genetics*, v. 83, p. 878-886, 1992b.

GODDARD, M. E. Genetic evaluation in the dairy industry. In: HAMMOND, K.; GRASER, H.V.; Mc DONALD, C.A. *Animal breeding: the modern approach*. Sydney: Post Graduate Foundation of Veterinary Science, 1992a. p. 95-101.

GOFFINET, B. Selection on selected records. *Genetique, Selection, Evolution*, v. 15, p. 91, 1983.

GOLDBERGER, A. S. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, Washington, v.57, p.369-375, 1962.

GOLUB, G. H.; LOAN, C. F. van. *Matrix computations*. Baltimore: John Hopkins University Press, 1983.

GOMES, M. I. Verossimilhança e inferencia estatística. In: COLÓQUIO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL, 2., 1981. *Actas*. Rio de Janeiro: UFRJ, 1981. p. 230-246.

GONÇALVES, P. S. Melhoramento genético da seringueira (*Hevea spp.*). In: SIMPÓSIO SOBRE A CULTURA DA SERINGUEIRA NO ESTADO DE SAO PAULO, 1., 1986, São Paulo. *Trabalhos apresentados...* Campinas: Fundacao Cargill, 1986. p. 95-123.

- GONÇALVES, P. S.; ORTOLANI, A. A.; CARDOSO, M. *Melhoramento genético da seringueira (Hevea spp.)* – uma revisão. Campinas: Instituto Agronômico de Campinas, 1995. (IAC. Documento, 54).
- GONÇALVES, P. S.; PAIVA, J. R.; SOUZA, R. A. *Retrospectiva e atualidade do melhoramento genético da seringueira (Hevea spp.) no Brasil e em países asiáticos*. Manaus: Embrapa-CNPSD, 1983. 51 p. (EMBRAPA-CNPSD. Documentos, 2).
- GONÇALVES, P.S.; CARDOSO, M.; SAES, L. A. Estimativas de repetibilidade na seleção de árvores de seringueira. *Pesquisa Agropecuária Brasileira*, Brasília, v. 25, n. 7, p. 1031-1038, 1990.
- GONZAGA NETO, L. Melhoramento genético da aceroleira na Embrapa Semi-Árido. In: QUEIRÓZ, M. A; GOEDERT, C. O; RAMOS, S. R. R. (Ed.) *Recursos genéticos e melhoramento de plantas para o Nordeste brasileiro*. Petrolina: EMBRAPA-CPATSA, 1999. Anais on line. (www.cpatsa.embrapa.br).
- GONZAGA NETO, L. Melhoramento genético da aceroleira. In: SÃO JOSÉ, A. R.; ALVES, R. E. (Ed.). *Acerola no Brasil, produção e mercado*. Vitória da Conquista: UESB, 1995. p. 15-21.
- GONZALEZ, V.; BREWBAKER, J. L.; HAMILL, D. E. Leucaena cytogenetics in relation to the breeding of low mimosine lines. *Crop Science*, v. 7, p. 140-143, 1967.
- 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.
- GRAYBILL, F. A. *Theory and application of the linear model*. North Scituate: Duxbury Press, , 1976. 704 p.
- GREEN, P. J.; SILVERMAN, B. W. *Nonparametric regression and generalized linear models*. London: Chapman & Hall. 1994. 182 p.
- GREGOIRE, T.G., SCHABENBERGER, O. Nonlinear mixed-effects modeling of cumulative bole volume with spatially correlated within-tree data. *Journal of Agricultural, Biological and Environmental Statistics*, v.1, n.1, p.107-119, 1996.
- GRIFFIN, A. R.; COTERILL, P. P. Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genetica*, v. 37, v. 3-4, p. 124-131, 1988.
- GRIFFIN, A. R.; MORAN, G. F.; FRIPP, Y. Preferential outcrossing in *Eucalyptus regnans*. *Australian Journal of Botany*, v. 35, p. 465-475, 1987.
- GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. *Australian Journal of Biological Sciences*, v. 9, p. 463-493, 1956.

GRIGNOLA, F. E.; HOESCHELE, I.; TIER, B. Mapping quantitative trait loci in outcross populations via residual maximum likelihood. I. Methodology. *Genetics, Selection, Evolution*, v. 28, p. 479-490, 1996.

GROENEVELD, E. VCE-A multivariate multimodel REML covariance component estimation program. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 5., 1994, Ontario. *Proceedings...* Genetic and breeding of dairy and beef cattle, swine and horses. Guelph: University of Guelph, 1994. v.22, p.46-48.

GROENEVELD, E.; KOVAC, M.; WANG, T. PEST, a general purpose BLUP package for multivariate prediction and estimation. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 4., 1990, Edinburgh. *Proceedings...* Edinburgh: [s.n.], 1990. v. 13, p. 488-491.

GRONDONA, M. O.; CROSSA, J.; FOX, P. N.; PFEIFFER, W. H. Analysis of variety yield trials using two-dimensional separable ARIMA processes. *Biometrics*, v. 52, p. 763-770, 1996.

HAAPANEN, M. Effect of plot size and shape on the efficiency of progeny tests. *Silva Fennica*, v. 26, n. 4, p. 201-209, 1992.

HALEY, C. S.; CAMERON, N. D.; SLEE, J.; LAND, R. B. Selección indirecta. In: MARAI, I. F. M.; OWEN, J. B. (Ed.). *Nuevas técnicas de producción ovina*. Acribia: Zaragoza, 1994. p. 123-135.

HALLAUER, A. R. Heterosis. In: COORS, J. G.; PANDEY, S. (Ed.). *The genetics and exploitation of heterosis in crops*. Madson: ASA-CSSA-SSSA, 1999. p. 481-489.

HAMMOND, K.; GRASER, H. V.; McDONALD, C. A. *Animal breeding*: the modern approach. Sydney: Post Graduate Foundation of Veterinary Science, 1992. p.257

HAMRICK, J. L. Distribution of genetic diversity in tropical tree populations: implications for the conservation of genetic resources. In: LAMBETH, C.C.; DVORAK, W. (Ed.) *Resolving tropical forest resource concerns thorough tree improvement, gene conservation and domestication of new species*. Raleigh: NC State University Press, 1992. p. 74-82.

HARDIYANTO, E. B. Approaches to breeding acacias for growth and form: the experience at Barito Pacifico Group. In: TURNBULL, H.R. et al. (Ed.) *Recent developments in Acacia planting*. Canberra: ACIAR, 1998. p. 178-183.

HARDNER, C.; BORRALHO, N. TIER, B.; MILLEK, S.; GODDARD, M. Accounting for dominance and inbreeding in genetic evaluations using individual tree mixed models. In: DIETERS, M. (Ed.). *Tree improvement for sustainable tropical forestry*. Caloundra: QFRI, 1996. p. 143-147.

HARRIS, J.A. On the calculation of the intraclass and interclass coefficients of correlation from class moments. *Biometrika*, v.9, p. 446-472, 1913.

- HARRIS, B.; JOHNSON, D. Approximate reliability of genetic evaluations under an animal model. *Journal of Dairy Science*, v. 81, p. 2723-2728, 1998.
- HARRIS, D. L. Expected and predicted progress from index selection involving estimates of population parameters. *Biometrics*, v. 20, p. 46-72, 1964.
- HARTL, L. D.; CLARK, A. G. *Principles of population genetics*. 2. ed. Sunderland Sinauer Associates, Inc. Publ., 1989. 681 p.
- HARTLEY, C. W. S. Oil palm selection and breeding. In: Hartley, C.W.S. *The oil palm*. London: Longman, 1977. p. 195-310.
- HARTLEY, H. O.; RAO, J. N. K. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, v. 54, p. 93-108, 1967.
- HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, Washington, v. 72, p. 320-328, 1977.
- HARVILLE, D. A.; MEE, R. W. A mixed-model procedure for analysing ordered categorical data. *Biometrics*, v. 40, p. 93-408, 1984.
- HARVILLE, D. A.; CARRIQUIRY, A. L. Classical and Bayesian prediction as applied to unbalanced mixed linear models. *Biometrics*, v. 48, p. 987-1003, 1992.
- HAUGERUD, A.; COLLINSON, M. P. Plants, genes and people: improving the relevance of plant breeding in Africa. *Experimental Agriculture*, v. 26, p. 341-362, 1990.
- HAZEL, L. N. The genetic basis for constructing selection indexes. *Genetics*, v. 28, p. 476-490, 1943.
- HAZEL, L.N.; LUSH, J.L. The efficiency of three methods of selection. *Journal of Heredity*, v. 33, p. 393-399, 1943.
- HAZEL, L.N.; TERRILL, C.E. Effects of some environmental factors on weaning traits of range Rambouillet lambs. *Journal of Animal Science*, v. 4, p. 331-341, 1945.
- HEDGE, N. Leucaena for energy plantation. *BAIF Journal*, v. 5, p. 37-42, 1985.
- HENDERSON JUNIOR, C. R. Analysis of covariance in the mixed model: higher-level, nonhomogeneous and random regression. *Biometrics*, v.38, p.633-640, 1982.
- HENDERSON JUNIOR, C. R.; HENDERSON, C. R. Analysis of covariance in mixed models with unequal subclass numbers. *Communications in Statistics*, v. A8, p. 751, 1979.
- HENDERSON, C. R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, v. 32, p. 69-83, 1976.

HENDERSON, C. R. *Applications of linear models in animal breeding*. Guelph: University of Guelph, 1984. 462 p.

HENDERSON, C. R. Avaliação de vacas e touros. In: SIMPÓSIO BRASILEIRO DE MELHORAMENTO GENÉTICO DE BOVINOS, 1., 1982. *Anais...* Coronel Pacheco: EMBRAPA-CNPGL, 1982a. p. 137-168.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, v. 31, p. 423-449, 1975.

HENDERSON, C. R. Best linear unbiased prediction in populations that have undergone selection. In: BARTON, R.A.; SMITH, W.C. (Ed.) *Proceedings of the World Congress on Sheep and Beef Cattle Breeding*. Palmerston North: Dunmore Press, 1982b, v. 1, p. 191-200.

HENDERSON, C. R. Best linear unbiased prediction of non-additive genetic merits in non-inbred populations. *Journal of Animal Science*, v. 60, p. 111-117, 1985.

HENDERSON, C. R. *Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine*. Ames: Iowa State University, 1948. Ph. Thesis.

HENDERSON, C. R. Estimation of changes in herd environment. *Journal of Dairy Science*, v. 32, p. 709, 1949.

HENDERSON, C. R. Estimation of genetic parameters. *Annals of Mathematical Statistics*, v. 21, p. 309, 1950.

HENDERSON, C. R. Estimation of variance and covariance components. *Biometrics*, v. 9, p. 226-252, 1953.

HENDERSON, C. R. Prediction of the merit of single crosses. *Theoretical and Applied Genetics*, v. 49, p.273, 1977.

HENDERSON, C. R. Prediction of future records. In: POLLAK, E.; KEMPTHORNE, O.; BAILEY, I. (Ed.). *Proceedings International Conference on Quantitative Genetics*. Ames: Iowa State University Press, 1977, p. 615-638.

HENDERSON, C. R. Sire evaluation and genetic trends. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF J. LUSH. Champaign, 1973. Champaign: American Society of Animal Science, 1973, p.10-41.

HENDERSON, C.R. Specific and general combining ability. In: GOWEN, J. W. *Heterosis*. New York: Hafner Publishing, 1952. p.352.

HENDERSON, C. R. Using estimates of variances in prediction of breeding values under a selection model. In: VLECK, L.D. van; SEARLE, S.R. (Ed.) *Variance components and animal breeding*. Ithaca: Cornell University, 1979. p. 217.

- HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R.; VON KROSIGH, C. M. The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, v. 15, p. 192, 1959.
- HENDERSON, C. R.; QUAAS, R. L. Multiple trait evaluation using relatives records. *Journal of Animal Science*, v. 3, p. 1188-1197, 1976.
- HENDERSON, C.R. Selection index and expected genetic advance. In: HANSON, W.D.; ROBINSON, H.F. (Ed.). *Statistical genetics and plant breeding*. Whashington: National Academy of Sciences, 1963. p. 141-163. (NAS-NCR Pub., 982).
- HENDERSON, C.R. Statistical methods in animal improvement: Historical overview. In: GIANOLA, D.; HAMMOND, K. (Ed.). *Advances in Statistical Methods for Genetic Improvement of Livestock*. Berlin: Springer-Verlag, 1990. p. 1-14.
- HIGA, A. R.; RESENDE, M. D. V. de. Breeding Acacia mearnsii in Southern Brazil. In: BROWN, A.G. (Ed.) *Australian tree species research in China*. Zhan-gzhou: ACIAR, 1992, p. 158-160. (ACIAR. Proceedings, 48).
- HIGA, A. R.; RESENDE, M. D. V. de; LAVORANTI, O. J. *Projeto melhoramento genético e conservação de Eucalyptus*. Colombo: EMBRAPA, 1993. 700 p. Relatório não publicado.
- HILL, W. G. A note on effective population size with overlapping generations. *Genetics*, v. 92, p. 317-322. 1979.
- HILL, W. G. Considerations in the design of animal breeding experiments. In: GIANOLA, D.; HAMMOND, K. (Ed.). *Advances in statistical methods for genetic improvement of livestock*. Berlin: Spring-Verlag, 1990. p. 59-76.
- HILL, W. G. Effective size of populations with overlapping generations. *Theoretical Population Biology*, v. 3, p. 278-289, 1972.
- HILL, W. G. Order statistics of correlated variables and implications in genetic selection programmes. *Biometrics*, v. 32, p. 889-902, 1976.
- HILL, W. G. Prediction and evaluation of response to selection with overlapping generations. *Animal Production*, v. 18, p. 117-140, 1974.
- HILL, W. G. Predictions of response to artificial selection from new mutations. *Genetical Research*, v. 40, p. 255-278, 1982.
- HILL, W. G. *Quantitative genetics*: Part 1. Explanation and analysis of quantitative variation. New York: Van Nostrand Reinhold Co., 1984. 550 p.
- HILL, W. G. Variation in response to selection. In: POLLAK, E.; KEMPTHORNE, O.; BAILEY, T. B. (Ed.). *Proceedings of the International Conference on Quantitative Genetics*. Ames, 1977. Ames: Iowa State University Press, 1977, p. 343-365.
- HILL, W. G.; CABALLERO, A. Artificial selection experiments. *Annual Review of Ecology and Systematics*, v. 23, p. 287-310, 1992.

HILL, W. G.; MACKAY, T. ***Evolution and animal breeding.*** Wallingford: CAB International, 1989. 313 p.

HILL, W. G.; NICHOLAS, F. W. Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment. ***Biometrics***, v. 30, p. 447-468, 1974.

HILL, W. G.; RABASH, J. Models of long-term artificial selection in finite populations. ***Genetical Research***, v. 48, p. 41-50, 1986.

HILL, W. G.; SMITH, C. Estimating heritability of a dichotomous trait. ***Biometrics***, v. 33, p. 234-236, 1977.

HINLKELMANN, K. Estimation of heritability from experiments with related dams. ***Biometrics***, v. 25, p. 755-766, 1969.

HINKELMAN, K.; KEMPTHORNE, O. ***Design and analysis of experiments:*** introduction to experimental designs. New York: John Wiley & Sons. 1994. 512 p.

HODGE, G. R.; VOLKER, P. W.; POTTS, B. M.; OWEN, J. V. A comparison of genetic information from open-pollinated and control – pollinated progeny tests in two eucalypt species. ***Theoretical and Applied Genetics***, v. 92, p. 53-63, 1996.

HODGE, G.; WHITE, T. L. Systematic designs for advanced-generation wind-pollinated seed orchard design. ***New Forests***, v.7, p.213-236, 1993.

HODGSON, L.M. Breeding of Eucalyptus in South Africa. ***South African Forestry Journal***, v. 89, p. 13-15, 1974.

HOESCHELE, I.; RADEN, P. M. van. Rapid inverse of dominance relationship matrices for noninbred populations by including sire-dam combination effects. ***Journal of Dairy Science***, v. 74, p. 570-579, 1991.

HOFER, A. Variance component estimation in animal breeding: a review. ***Journal of Animal Breeding and Genetics***, v. 115, p. 247-265, 1998.

HORNER, T. W.; WEBER, C. R. Theoretical and experimental study of self fertilized populations. ***Biometrics***, v. 12, p. 404-414, 1956.

HUBER, D. A.; WHITE, T. L.; HODGE, G. R. The efficiency of half-sib, half diallel and circular mating designs in the estimation of genetic parameters in forestry: a simulation. ***Forest Science***, v. 38, n. 4, p. 757-776, 1992.

HUGHES, C. E. New opportunities in Leucaena genetic improvement. In: GIBSON, G. L.; GRIFFIN, A. R.; MATHESON, A. C. (Ed.). ***Breeding tropical trees:*** population structure and genetic improvement strategies in clonal and seedling forestry. Oxford: Oxford Forestry Research Institute, 1989. p. 218-226.

HUTTON, E. M. Selection and breeding of Leucaena for very acid soils. In: LEUCAENA RESEARCH IN THE ASIA-PACIFIC REGION, Ottawa, 1983. ***Proceedings...*** Ottawa: IDRC, 1983. p.23-26.

- IBANEZ, M.; CARABANO, M. J.; FOULLEY, J. L.; ALENDÁ, R. Heterogeneity of herd-period phenotypic variances in the Spanish Holstein-Friesian cattle: sources of heterogeneity and genetic evaluation. *Livestock Production Science*, v. 45, p. 137-147, 1996.
- IEMMA, A. F. *Modelos lineares* – uma introdução para profissionais da pesquisa agropecuária. Londrina: UEL / RBRAS, 1987. 262 p.
- JACKSON, N.; TURNER, H. N. Optimal structure for a cooperative breeding system. *Proceedings of the Australian Society for Animal Production*, v. 9, p. 55-69, 1972.
- JACQUARD, A. Inbreeding: one word, several meanings. *Theoretical Population Biology*, v. 7, p. 338-363, 1975.
- JAMES, J. W. Design of animal breeding programs. In: HILL, W.G.; MACKAY, T. (Ed.). *Evolution and animal breeding*. Wallingford; CAB International, 1989. p. 189-194.
- JAMES, J. W. Effective population size in open nucleus breeding schemes. *Acta Agriculturae Scandinavica*, v. 28, p. 387-392, 1978.
- JAMES, J. W. Index selection with restrictions. *Biometrics*, v. 24, p. 1115-1018, 1968.
- JAMES, J. W. Open nucleus breeding system. *Animal Production*, v. 24, p. 287-303, 1977.
- JAMES, J. W. The founder effect and response to artificial selection. *Genetical Research*, v. 16, p. 41-250, 1971.
- JAMES, W.; STEIN, C. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, v. 1, p. 361-379, 1961.
- JAMROZIK, J.; SCHAEFFER, L. R. Estimates of genetic parameters for a test day model with random regressions for yield of first lactation Holsteins. *Journal of Dairy Science*, v. 80, p. 726-770, 1997.
- JAMROZIK, J.; SCHAEFFER, L. R.; DEKKERS, J. C. M. Genetic evaluation of dairy cattle using test day yields and random regression model. *Journal of Dairy Science*, v. 80, p. 1217-1226, 1997.
- JARVIS, S. F.; BORRALHO, N. M. G.; POTTS, B. M. Implementation of a multi-variate BLUP model for genetic evaluation of *E. globulus* in Australia. In: POTTS, B; VOLKER,P.; BORRALHO, N.M.G.. (Ed.). *Eucalypt plantations: improving fiber yield and quality*. Hobart: CRC / IUFRO, 1995. p. 212-216.
- JEFFREYS, H. *Theory of probability*. Oxford : Clarendon Press, 1961. 459 p.
- JENKINS, M. T. Methods of estimating the performance of double crosses in corn. *Journal of the Americann Society of Agronomy*, v. 26, p. 199-204, 1934.

JENSEN, J.; WANG, C. S. ; SORENSEN, D. A.; GIANOLA, D. Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects using the Gibbs sampler. *Acta Agriculturae Scandinavica*, v. 44, p. 192-201, 1994.

JOHANSSON, K.; KENNEDY, B.; QUINTON, M. Prediction of breeding values and dominance effects from mixed models with approximations of the dominance relationship matrix. *Livestock Production Science*, v. 34, p. 213-223, 1993.

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.

JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. Englewood : Prentice Hall Inc., 1988. 594 p.

JONES, L. P.; FRANKHAM; BARKER, J. S. F. The effects of population size and selection intensity in selection for a quantitative character in *Drosophila* III. Long-term response to selection. *Genetical Research*, v. 12, p. 249-266, 1968.

KACKAR, R. N.; HARVILLE, D. A. Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics*, v. A10, p. 1249, 1981.

KAGEYAMA, P. Y. Genetic structure of tropical tree species of Brazil. In: BAWA, K.S.; HADLEY, M. (Ed.). *Reproductive ecology of tropical forest plants*. Paris: Unesco, 1990. p.375-387.

KAGEYAMA, P. Y. *Seleção precoce a diferentes idades de progêneres de Eucalyptus grandis (Hill) Maiden*. Piracicaba: ESALQ, 1983, 147 p. Tese Livre Docência.

KAGEYAMA, P. Y. Conservação "in situ" de recursos genéticos de plantas. *IPEF*, Piracicaba, v. 35, p. 7-37, 1987.

KAGEYAMA, P. Y. *Variação genética em uma população de Eucalyptus grandis*. Piracicaba: ESALQ, 1980. 125 p. Tese Doutorado.

KAGEYAMA, P. Y.; DIAS, I. S. Aplicación de conceptos genéticos a especies forestales nativas en Brasil. *Información sobre Recursos Genéticos Forestales*, v. 13, p. 2-10. 1985.

KAGEYAMA, P. Y.; MORA, A. L.; BARRICHELO, L. E. G.; MIGLIORINI, A. J.; SANSIGOLO, C. A. Variação genética para densidade da madeira em progêneres de *Eucalyptus grandis*. *Silvicultura*, v. 8, n. 28, p. 318-324, 1983a.

KAGEYAMA, P. Y.; PINTO JUNIOR, J. E.; MORA, A. L.; NICOLIELO, N. Half-sib progeny trial of *Pinus caribaea* var. *hondurensis* from selected superior trees of Australian populations. *Silvicultura*, São Paulo, v. 8, p. 97-99, 1983b.

- KAGEYAMA, P. Y.; VENCOVSKY, R. Variação genética em progêneres de uma população de *Eucalyptus grandis* Hill Maiden. *IPEF*, Piracicaba, v. 24, p. 9-26, 1983.
- KANG, H. Designing a tree breeding system. In: SEVENTEENTH MEETING CANADIAN TREE IMPROVEMENT ASSOCIATION. Truro, 1979. Truro: CTIA, 1979. p. 51-66.
- KANG, H; NAMKOONG, G. Limits of artificial selection under unbalanced mating systems. *Theoretical and Applied Genetics*, v. 58, p. 181-191, 1980.
- KEEN, A.; ENGEL, B. Analysis of a mixed model for ordinal data by iterative re-weighted REML. *Statistica Neerlandica*, v. 51, n. 2, p. 129-144, 1997.
- KEMPTHORNE, O. *Design and analysis of experiments*. New York: J. Wiley, 1952. 631 p.
- KEMPTHORNE, O. *An introduction to genetic statistics*. New York: J. Wiley, 1957. 545 p.
- KEMPTHORNE, O.; NORDSKOG, A. W. Restricted selection indices. *Biometrics*, v. 15, p. 10-19, 1959.
- KEMPTHORNE, O.; CURNOW, R. N. The partial diallel cross. *Biometrics*, v. 17, p. 229-250, 1961.
- KENNEDY, B. W. Use of mixed model methodology in analysis of designed experiments. In: GIANOLA, D.; HAMMOND, K. (Ed.). *Advances in statistical methods for genetic improvement of livestock*. Berlin: Springer-Verlag, 1990. p. 77-97.
- KENNEDY, B. W.; TRUS, D. Considerations on genetic connectedness between management units under an animal model. *Journal of Animal Science*, v. 71, p. 2341-2352, 1993.
- KENNEDY, B. W. Charles Roy Henderson: The unfinished legacy. *Journal of Dairy Science*, v. 74, n. 11, p. 4067-4081, 1991.
- KENNEDY, B. W.; SORENSEN, D. A. Properties of mixed model methods for prediction of genetic merit under different genetic models in selected and unselected populations. In: WEIR, B.: GOODMAN, M.M.; NAMKOONG, G. (Ed.). *Second International Conference on Quantitative Genetics*, Raleigh, 1988. Raleigh: North Carolina State University, 1988. p. 91-103.
- KENWARD, M. G.; ROGER, J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, v. 53, n. 3, p. 983-997, 1997.
- KERR, R. J.; GODDARD, M. E. Comparisons between the use of MAS and clonal tests in tree breeding programmes. In: IUFRO'97 GENETICS OF RADIATA PINE, 1997, Rotorua. *Proceedings...* Rotorua: NZFRI, 1997. p. 297-303. (FRI Bulletin, 203).

KEULS, M. The use of the studentized range in connection with an analysis of variance. *Euphytica*, v. 1, p. 112-122, 1952.

KEULS, M.; GARRETSEN, F. A general method for the analysis of genetic variation in complete and incomplete diallels and North Carolina II designs. Part 1 : Procedures and general formulas for the random model. *Euphytica*, v. 26, p. 537-551, 1977.

KHA, L. D.; HAI, N. D.; VINH, H. Q. Clonal tests and propagation options for natural hybrids between *Acacia mangium* and *A. auriculiformis*. In: TURNBULL, H.R.; CROMPTON, H. R.; PINYOPUSARERK, K. (Ed.) *Recent developments in Acacia planting*. Canberra: ACIAR, 1998. p. 203-210.

KHURI, A. I.; LITTELL, R. C. Exact tests for the main effects variance components in an unbalanced random two-way model. *Biometrics*, v. 43, n. 2, p. 545-560, 1987.

KIKUTI, P. *Parâmetros genéticos em progênieis de meios irmãos e clonais numa população de Eucalyptus grandis na região de Telêmaco Borba*. Piracicaba: ESALQ, 1988. 119 p. Tese Mestrado.

KING, J. N.; JOHNSON, G. R. Monte Carlo simulation models of breeding population advancement. *Silvae Genetica*, v. 42, n. 2/3, p. 68-78, 1993.

KINGHORN, B. Principles of estimated breeding values. In: HAMMOND, R.; GRASER, H-U.; MC DONALD, C.A. *Animal breeding – the modern approach*. Sydney: Post Graduate Foundation in Veterinary Science, 1992. p. 47-55.

KINGHORN, B. P. Crossbreeding in domestic animals. *Proceedings of the Australian Association of Animal Breeding and Genetics*, v. 6, p. 112, 1987.

KIRKPATRICK, M.; HILL, W. G.; THOMPSON, R. Estimating the covariance structure of traits during growth and ageing, illustrated with lactations in dairy cattle. *Genetical Research*, v. 64, p. 57-69, 1994.

KIRKPATRICK, M.; LOFSVOLD, D.; BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*, v. 124, p. 979-993, 1990.

KNUIMAN, M. W.; LAIRD, N. M. Parameter estimation in variance component models for binary response data. In: GIANOLA, D.; HAMMOND, K. (Ed.). *Advances in statistical methods for genetic improvement of livestock*. Berlin: Springer-Verlag, 1990. p. 177-206.

KOCH, R. M.; CLARK, R. T. Genetic and environmental relationships among economic characters in beef cattle. III. Evaluating maternal environment. *Journal of Animal Science*, v. 14, p. 979-996, 1955.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one criterion variance analysis. *Journal of the American Statistical Association*, v. 47, p. 583-621, 1952.

- KUSNADAR, D.; GALWEY, N. A proposed method for estimation of genetic parameters on forest trees without raising progeny: critical evaluation and refinement. *Silvae Genetica*, v. 49, p. 15-21, 2000.
- LAMBETH, C; GLADSTONE, W.; STONECYPHER, R. Statistical efficiency of row and noncontiguous family plots in genetic tests of loblolly pine. *Silvae Genetica*, v.32, n. 1-2, p. 24-8, 1983.
- LAMKEY, K. R. Discussion session. In: COORS, J.G.; PANDEY, S. (Ed.). **The genetics and exploitation of heterosis in crops**. Madison: ASA / CSSA / SSSA, 1999. p. 94-98.
- LAMKEY, K. R.; EDWARDS, F. Quantitative genetics of heterosis. In: COORS, J.G.; PANDEY, S. (Ed.). **The genetics and exploitation of heterosis in crops**. Madison: ASA / CSSA / SSSA, 1999. p. 29-43.
- LANDE, R. Mutation and conservation. *Conservation Biology*, v. 9, p. 782-791, 1995.
- LANDE, R. Risk of populations extinction from fixation of new deleterious mutations. *Evolution*, v. 48, p.1460-1469, 1994.
- LANDE, R. Genetics and demography in biological conservation. *Science*, v. 241, p. 1455-1460. 1988.
- LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, v. 124, p. 743-756, 1990.
- LANDIM, P. M. B. *Análise estatística de dados geológicos*. São Paulo: Fundação Ed. da UNESP, 1998. 226 p.
- LANGE, K.; SINSHEIMER, J. S. Normal/independent distributions and their applications in robust regression. *Journal of the American Statistical Association*, v. 2, p. 175-198, 1993.
- LATTER, B. D. H.; MULLEY, J. C.; REID, D.; PASCOE, L. Reduced genetic load revealed by slow inbreeding in *Drosophila melanogaster*. *Genetics*, v. 139, p. 287-297, 1995.
- LATTER, B. D. H.; ROBERTSON, A. Experimental design in the estimation of heritability by regression methods. *Biometrics*, v. 16, p. 348-353, 1960.
- LEBART, L.; MORINEAU, A.; FENELON, J. P. *Treatments des données statistiques*. Paris: Dunod, 1979. 510 p.
- LECORRE, V.; ROUSSEL, G.; ZANETTO, A.; KREMER, A. Geographical structure of gene diversity in *Quercus petraea*. III. Patterns of variation identified by geostatistical analysis. *Heredity*, v. 80, p. 464-473, 1998.
- LECOUSTRE, R.; REFFYE, P. La théorie des variables régionalisées, des applications possibles dans le domaine épidémiologique aux recherches agronomiques, en particular sur le palmier à huile et le cocotier. *Oléagineux*, v. 41, n. 12, p. 541-548, 1986.

LEDIG, F. T.; WHITMORE, J. L. Heritability and genetic correlations for volume, foxtails and other characteristics of caribbean pine in Puerto Rico. *Silvae Genetica*, v. 30, n. 2-3, p. 88-92, 1981.

LEE, C. H. Statistical efficiency varies with plot size, number of replications and seedlots sampled, *Silvae Genetica*, v. 32, p. 49-52, 1983

LERNER, I. M. *The genetic basis of selection*. New York: J. Wiley & Sons, 1958. 298 p.

LERNER, I. M.; HAZEL, L. N. Population genetics of a poultry flock under artificial selection. *Genetics*, v. 32, p. 325-339, 1947.

LEROY, T.; MONTAGNON, C.; CILAS, C.; CHARRIER, A.; ESKES, A.B. Reciprocal recurrent selection applied to *Coffea canephora*, 2. Estimation of genetic parameters. *Euphytica*, v. 74, n. 1-2, p. 121-128, 1994.

LI, C. C. *First course in population genetics*. Pacific Grove: Boxwood Press, 1955. 480 p.

LIBBY, W. J.; COCKERHAM, C. C. Randonn non-contiguous plots in interlocking field layouts. *Silvae Genetica*, v. 29, p. 183-190, 1980.

LIN, C. S.; BINNS, M. R. Working rules for determining the plot size and number of plots per block in field experiments. *Journal of Agricultural Sciences*, v. 103, p. 11-15, 1984.

LINDGREN, D.; GEA, L.; JEFFERSON, P. Loss of genetic diversity monitored by status number. *Silvae Genetica*, v. 45, n. 1, p. 52-59, 1996.

LINDLEY, D. V. *Introduction to probability and statistics*. Cambridge: Cambridge University Press, 1965. 188 p.

LINDLEY, D. V. The bayesian approach. *Scandinavian Journal of Statistics*, v. 5, p. 1-26, 1978.

LINDLEY, D. V.; SMITH, A. F. M. Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B*, v. 34, p. 1-41, 1972.

LINDSEY, J. K. *Applying generalized linear models*. New York: Springer-Verlag, 1997. 256 p.

LINDSEY, J. K. Some statistical heresies. *Journal of the Royal Statistical Society: Series B*, 1999.

LISBÃO JUNIOR, L. Avaliação e interpretação da interação genótipo x ambiente. I. Estabilidade fenotípica. In: CONGRESSO FLORESTAL E DO MEIO AMBIENTE DO PARANÁ, 3., 1991. *Anais...* Curitiba: Associação Paranaense de Engenheiros Florestais, 1991. p. 189-201.

LOESCHKE, V.; TOMIUK, J.; JAIN, S.K. Introductory remarks: genetics and conservation biology. In: LOESCHKE, V.; TOMIUK, J.; JAIN, S. K. *Conservation Genetics*. Birkhäuser, Verlag, Basel. 1994. p. 3-8.

- LONGFORD, N. T. *Random coefficients models*. Oxford: Clarendon Press, 1993. 390 p. (Oxford Statistical Science Series).
- LOO-DINKS, J. A.; TAUER, C. G. Statistical efficiency of six progeny test field designs on three loblolly pine (*Pinus taeda* L.) site types. *Canadian Journal of Forest Research*, v. 17, p. 1066-1070, 1987.
- LOPEZ-FANJUL, C. Tests of theory by selection experiments. In: HILL, W. G.; MACKAY, T. F. C. *Evolution and animal breeding*. Wallingford: CAB International, 1989, p. 129-133.
- LOTODÉ, R. Possibilités d'amélioration de l'expérimentation sur cacaoyers. *Café Cacao Thé*, n. 15, p. 91-104, 1971.
- LOTODÉ, R.; LACHENAUD, P. Méthodologie destinée aux essais de sélection du cacaoyer. *Café Cacao Thé*, v. 32, n. 4, p. 275-292, 1988.
- LOWE, W. J.; BUIJTENEN, J. P. van. The development of a sublining system in an operational tree improvement program. In: IUFRO CONFERENCE ON BREEDING THEORY, PROGENY TESTING AND SEED ORCHARDS. Williamsburg, 1986. Williamsburg: IUFRO, p.98-106.
- LUSH, J. L. *Animal breeding plans*. 3. ed. Ames: Iowa State University Press, 1945. 443p.
- LUSH, J. L. *Animal breeding plans*. Ames: Iowa State University Press. 1937. 433p.
- LUSH, J. L. Chance as a cause of gene frequency within pure breeds of livestock. *American Naturalist*, v. 80, p. 318-342, 1946.
- LUSH, J. L. Genetics aspects of the Danish system of progeny testing swine. *Iowa Research Bulletin*, n. 204, 1936. 17p.
- LUSH, J. L. Repeatability of threshold traits. *Biometrics*, v. 12, p. 84-88, 1956.
- LUSH, J. L. The number of daughters necessary to prove a sire. *Journal of Dairy Science*, v. 14, p. 209-220, 1931.
- LUSH, J. L. Family merit and individual merit as bases for selection. *American Naturalist*, v. 81, p. 241-261, 1947.
- LUSH, J. L. Progeny test and individual performance as indicators of an animal's breeding value. *Journal of Dairy Science*, v. 18, p. 1-19, 1935.
- LUSH, J. L.; LAMOREUX, W. F.; HAZEL, L. N. The heritability of resistance to death in the fowl. *Poultry Science*, v. 27, p. 375-388, 1948.
- LYNCH, M.; WALSH, B. *Genetics and analysis of quantitative traits*. Sunderland: Sinauer Associates, Inc., 1997. 980 p.
- MADALENA, F. E.; HILL, W. G. Population structure in artificial selection programmes: simulations studies. *Genetical Research*, v. 20, p. 75-99, 1972.

MADALENA, F. E.; ROBERTSON, A. Population structure in artificial selection: studies with *Drosophila melanogaster*. *Genetical Research*, v. 24, p. 113-126, 1975.

MAGNABOSCO, C. U. *Estimativas de parâmetros genéticos em características de crescimento de animais da raça nelore usando os métodos REML e amostragem de Gibbs*. Ribeirão Preto: USP / FMRP. 1997. 83 p. Tese Doutorado.

MAHALOVICH, M. F.; BRIDGEWATER, F. E. Modeling elite populations and positive assortative mating in recurrent selection programs for general combining ability. In: TWENTIETH SOUTHERN FOREST TREE IMPROVEMENT CONFERENCE. Charleston, 1989. *Proceedings...* Charleston: SFTIC, 1989. p.43-49.

MALHEIROS, E. B. *Efeitos da recuperação da informação interblocos na inferência estatística em ensaios de blocos incompletos equilibrados*. Piracicaba: ESALQ, 1982. Tese Doutorado.

MALHEIROS, E. B.; GODOI, C. R. M. Efeitos da estimação do parâmetro $a = w/w$ no teste F para a análise de variância, quando se faz a recuperação da informação interblocos. *Revista de Matemática e Estatística*, São Paulo, v. 2, p. 69-75, 1984b.

MALHEIROS, E. B.; GODOI, C. R. M. Estudo da distribuição do estimador do parâmetro $a = w/w$ na análise com a recuperação da informação interblocos. *Revista de Matemática e Estatística*, São Paulo, v. 2, p. 59-68, 1984a.

MANDEL, J. A new analysis of variance model for non-additive data. *Tecnometrics*, v. 13, n. 1, p. 1-18, 1971.

MANDEL, J. Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association*, v. 56, p. 878-888, 1961.

MANLY, B. F. J. *Randomization, bootstrap and Monte Carlo methods in biology*. 2. ed. London : Chapman and Hall, 1997. 399 p.

MANOJ, P. S.; GEORGE, T. E. Heterosis in cashew. *Cashew*, v. 7, n. 3, p.7- 9, 1993.

MARTIN, B. The benefits of hybridization: how do you breed for it? In: IUFRO CONFERENCE ON BREEDING TROPICAL TREES, 1989, Pattaya. *Proceedings...* Oxford: Oxford Forestry Institute, 1989. p. 79-92.

MARTIN, R. J. A subclass of lattice processes applied to a problem in planar sampling. *Biometrika*, v. 66, p. 209-217, 1979.

MARTIN, R. J. On the design of experiments under spatial correlation. *Biometrika*, v. 73, p. 247-277, 1986.

MARTIN, R. J. Some aspects of experimental design and analysis when errors are correlated. *Biometrika*, v. 69, p. 597-612, 1982.

- MARTIN, R. J. The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statistics: Theory and Methods*, v. 19, n. 1, p. 55-81, 1990.
- MARTINEZ, R. Control de la correlación espacial en experimentos de campo en el sector agrícola. *Agronomía Colombiana*, v. 11, p. 83-89, 1994.
- MARTINS, P. S. Estrutura populacional, fluxo gênico e conservação "in situ". *IPERF*, v. 35, p. 71-78, 1987.
- MATHER, K.; JINKS, J.L. *Introduction to biometrical genetics*. London: Chapman & Hall, 1977. p. 242
- MATHESON, A. C. Statistical methods and problems in testing large numbers of genotypes across sites. In: IUFRO CONFERENCE ON BREEDING TROPICAL TREES, Pattaya, 1989. *Proceedings...* Oxford: Oxford Forestry Institute, 1989, p. 93-105.
- MATHESON, A. C.; RAYMOND, C. A. Effects of thinning in progeny tests on estimates of genetic parameters in *Pinus radiata*. *Silvae Genetica*, v. 4-5, p. 125-128, 1984.
- MATOS, C. A. P.; THOMAS, D. L.; GIANOLA, D.; TEMPELMAN, R. J. Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models I. Estimation of genetic parameters. *Journal of Animal Science*, v. 75, p. 76-87, 1997.
- MATSUOKA, S.; GARCIA, A. A. F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A. (Ed.). *Melhoramento de espécies cultivadas*. Viçosa: Ed. da UFV, 1999. p. 205-251.
- MC CULLAGH, P.; NELDER, J. A. *Generalized linear models*. 2. ed. London: Chapman and Hall, 1989. 511 p.
- MC GUIRK, B. J. The estimation of genetic parameters for all-or-none and categorical traits. In: HILL, W.G.; MACKAY, T. (Ed.). *Evolution and animal breeding*. Oxford: CAB International, 1989. p. 175-180.
- MCCUTCHAN, B. G.; NAMKOONG, G; GIESBRECHT, F. G. Designs efficiencies with planned and unplanned unbalance for estimating heritability in forestry. *Forest Science*, v. 35, p. 801-815, 1989.
- McKEAND, S.; BEINEKE, F. Sublining for half-sib breeding populations of forest trees. *Silvae Genetica*, v. 29, p. 14-17, 1980.
- McKEAND, S.; BRIDGWATER, F. E. Third-generation breeding strategy for the North-Carolina State University - Industry Cooperative Tree Improvement Program. In: IUFRO Conference on Breeding Tropical Trees. Cali, 1992. *Proceedings...* Raleigh: North Carolina State University, 1992. p.234-240.

MCRAE, S.; COTTERILL, P. P. Macropropagation and micropropagation of *Eucalyptus globulus*: means of capturing genetic gain. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCALYPT, 1997, Salvador. **Proceedings...** Colombo: Embrapa Florestas, 1997, v.2, p. 14-17.

MENDES, A. N. Métodos de melhoramento aplicados na cultura do cafeiro. In: NUMBERG, P.L. et al. (Ed.). Simpósio de Atualização em Genética e Melhoramento de Plantas, 3, 1999, Lavras. **Anais...** Lavras: UFLA, 1999. p.18-35.

MEUNIER, J.; GASCON, J. P. Le schema général d'amélioration du palmier à huile à IRHO. **Oléagineux**, v. 27, p. 1-12, 1972.

MEYER, K. Approximate accuracy of genetic evaluation under an animal model. **Livestock Production Science**, v. 21, p. 87-100, 1989.

MEYER, K. DFREML – a set of programs to estimate variance components under an individual animal model. **Journal of Dairy Science**, Champaign, n. 2, Suppl., p. 33-34, 1988.

MEYER, K. DFREML – Version 3.0β: user Notes. Armidale: Institute of Animal Genetics of Edinburgh / Animal Genetics and Breeding Unit, University of New England, 1998. 31 p.

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. Estimation of genetic parameters. In: HILL, W.G.; MACKAY, T. F. C. **Evolution and animal breeding**. Wallingford: CAB International, 1989a. p. 161-167.

MEYER, K. Parameter estimation with covariance functions and random regression models. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. **Proceedings...** Armidale: AGBU/University of New England, 1998. v.25, p.517.

MEYER, K. Restricted maximum likelihood to estimate variance components for mixed models with two random factors. **Genetique, Selection, Evolution**, Paris, v. 19, p. 49-68, 1987.

MEYER, K. Restricted maximum likelihood to estimate variance componentes for animal models with several random effects using a derivative-free algorithm. **Genetique, Selection, Evolution**, Paris, v. 21, p. 317-340, 1989b.

MEYER, K. “DXMRR” – A program to estimate covariance functions for longitudinal data by restricted maximum likelihood. In: WORLD CONGRESS ON GENETIC APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. **Proceedings...** Armidale: AGBU/University of New England, 1998b. v. 27, p. 465-466.

MEYER, K. An average information restricted maximum likelihood algorithm for

estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices. *Genetics, Selection, Evolution*, v. 29, p. 97-116, 1997.

MEYER, K. Estimating covariance functions for longitudinal data using a random regression model, *Genetics, Selection, Evolution*, v. 30, p. 221-249, 1998a.

MEYER, K.; HILL, W. G. *Estimation of genetic and phenotypic covariance functions for longitudinal or repeated records by restricted maximum likelihood*. *Livestock Production Science*, v. 47, p. 185-200, 1997.

MILLER, R. G. The Jackknife – a review. *Biometrika*, v. 61, p. 1-15, 1974.

MILLER, S. P.; GODDARD, M. E. *A model of dominance and epistatic variation between and within breeds*. [S.l.: s.n.], 1997b. não publicado.

MILLER, S. P.; GODDARD, M. E. *Prediction of additive and dominance genetic effects (within and between breeds) in a population of breeds and crosses*. [S.l.: s.n.], 1997. não publicado.

MILLIKEN, G. A. ; JOHNSON, D. E. *Analysis of messy data I*: designed experiments. Belmont: Lifetime Learning Pub., 1984. 473 p.

MIRANDA FILHO, J. B. Princípios de experimentação e análise estatística. In: PATERNIANI, E. (Ed.). *Melhoramento e produção de milho no Brasil*. Piracicaba: Fundação Cargill, 1987, v. 2, p. 765-795.

MISZTAL, I. Comparison of software packages in animal breeding. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 5., 1994, Guelph. *Proceedings...* Guelph: University of Guelph, 1994, v.22, p.3-10.

MISZTAL, I. Estimation of variance components with large scale dominance models. *Journal of Dairy Science*, v. 80, p. 965-974, 1997.

MISZTAL, I.; GIANOLA, D. Indirect solution of mixed model equations. *Journal of Dairy Science*, v. 70, p. 716-723, 1987.

MOLENBERGHS, G.; FITZMAURICE, G. M.; LIPSITZ, S. R. Efficient estimation of the intraclass correlation for a binary trait. *Journal of Agricultural, Biological and Environmental Statistics*, v. 1, n. 1, p. 78-96, 1996.

MONTAGNON, C.; GUYOT, B.; CILAS, C.; LEROY, T. Genetic parameters of several biochemical compounds from green coffee, *Coffea canephora*. *Plant Breeding*, v. 117, n. 6, p. 576-578, 1998.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. Introduction to the theory of statistics. Tokyo: McGraw-Hill, 1974. 564 p.

MORAES, M.L.T. *Variação genética da densidade básica da madeira em progêniess de Eucalyptus grandis e suas relações com as características de crescimento*. Piracicaba: ESALQ, 1987. 115 p. Tese Mestrado.

MORAN, G. F.; BELL, C. Eucalyptus. In: TANKSLEY, S.D.; ORION, T.Y. *Iso-*

zymes in plant genetics and breeding, Part B. Amsterdam: Elsevier, 1983. p. 423-441.

MORAN, G. F.; BROWN, A. H. D. Temporal heterogeneity of outcrossing rates in alpine ash (*Eucalyptus delegatensis*). **Theoretical and Applied Genetics**, v. 57, p. 101-105, 1980.

MORETTIN, P. A.; TOLOI, C. M. **Previsão de séries temporais**. 2. ed. São Paulo: Atual, 1987. 439 p.

MORI, E. S.; KAGEYAMA, P. Y.; FERREIRA, M. Variação genética e interação progénie x locais em *Eucalyptus urophylla*. **IPEF**, Piracicaba, v. 39, p. 53-63, 1988.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. Wallingford: CAB International, 1996. 187 p.

MUIR, W. M. Candidate gene selection. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 34, 1997, Juiz de Fora. **Anais do Simpósio “Tópicos Especiais em Zootecnia”**. Juiz de Fora: Embrapa Gado de Leite Juiz de Fora, 1997, p.79-91.

MUNIZ, C. A. S. D.; CARVALHEIRO, R.; QUEIRÓZ, S. A.; FRIES, L. A. Critérios de seleção baseados em médias aritméticas e médias harmônicas. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 38., 2001, Piracicaba. **Anais...** Piracicaba: FEALQ, 2001. CD ROM.

MUONA, O.; HARJU, A. Effective population sizes, genetic variability, and mating system in natural stands and seed orchards of *Pinus sylvestris*. **Silvae Genetica**, Frankfurt, v. 38, n. 5-6, p. 221-228, 1989.

MURTEIRA, B. J. F. **Probabilidade e estatística**: inferência estatística. 2. ed. Lisboa: McGraw-Hill, 1990. v. 2. 480 p.

MURTEIRA, B. J. F. **Estatística**: inferência e decisão. Lisboa: Imprensa Nacional, Casa da Moeda, 1988. 380 p.

NAGYLAKI, T. **Introduction to theoretical population genetics**. Berlin: Springer-Verlag, 1992. 350 p.

NAMKOONG, G. A gene conservation plan for loblolly pine. **Canadian Journal of Forest Research**, v. 27, p. 433-437. 1997.

NAMKOONG, G. Inbreeding effects on estimation of genetic additive variance. **Forest Science**, v. 12, p. 8-13, 1966.

NAMKOONG, G. **Introduction to quantitative genetics in forestry**. Washington: USDA, 1979, 329 p.

NAMKOONG, G. A multiple index selection strategy. **Silvae Genetica**, v. 25, p. 199-201, 1976.

NAMKOONG, G.; BARNES, R. D.; BURLEY, J. **A philosophy of breeding**

- strategy for tropical forest tree.** Oxford: Commonwealth Forestry Institute, 1980. 67 p. (Tropical Forestry Paper, 16).
- NANSON, A. The provenance seedling seed orchards. *Silvae Genetica*, v. 21, p. 243-249, 1972.
- NEIMANN-SORENSEN, A.; ROBERTSON, A. The association between blood groups and several production characters in three Danish Cattle breeds. *Acta Agriculturae Scandinavica*, v. 11, p. 163-196, 1961.
- NELDER, J. A.; MEAD, R. A simplex method for function minimization. *Computer Journal*, v. 7, p. 147-151, 1965.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society: Series A*, v. 135, p. 370-384, 1972.
- NETER, J.; KUTNER, M. H.; NACHTSHEIN, C. J.; WASSERMAN, W. *Applied linear statistical models*. Chicago: Irwin, 1996. 1408 p.
- NEWMAN, D. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika*, v. 31, p. 20-30, 1939.
- NEWMAN, D.; PILSON, D. Increased probability of extinction due to decrease genetic effective population size: experimental populations of Clarkia pulchella. *Evolution*, v. 51, n. 2, p. 354-362. 1997.
- NICHOLAS, F. W. *Introduction to veterinary genetics*. New York: Oxford University Press, 1996. 317 p.
- NICHOLAS, F. W. Size of population required for artificial selection. *Genetical Research*, v. 35, p. 85-105, 1980.
- NICHOLAS, F. W. *Veterinary genetics*. Oxford: Clarendon Press, 1987. 580 p.
- NIETO, B.; SALGADO, C.; TORO, M. A. Optimization of artificial selection response. *Journal of Animal Breeding and Genetics*, v. 103, p. 199-204, 1986.
- NOITON, D. A. M.; SHELBOURNE, C. J. A. Quantitative genetics in an apple breeding strategy. *Euphytica*, v. 60, p. 213-219, 1992.
- NORDSKOG, A. W. Success and failure of quantitative genetic theory in poultry. In: POLLAK, E.; KEMPTHORNE, O.; BAILEY, T.B. (Ed.). *Proceedings of the International Conference on Quantitative Genetics*. Ames: Iowa State University Press, 1977, p.569-586.
- ODA, S.; MENCK, A. L. M.; VENCOVSKY, R. Problemas no melhoramento clássico de eucalipto em função da alta intensidade de seleção. *IPEF*, Piracicaba, v. 41/42, p. 8-17, 1989.
- OLAUSSON, A.; RONNINGEN, K. The estimation of genetic parameters for threshold characters. *Acta Agri. Scand.*, v. 26, p. 201-208, 1975.

OLIVEIRA, J. R. P.; SOARES FILHO, W. dos S. Acerola: conservação, caracterização e seleção de germoplasma pelo CNPMF-EMBRAPA. In: SÃO JOSÉ, A.R.; ALVES, R.E. (Ed.). **Acerola no Brasil, produção e mercado.** Vitória da Conquista: UESB, 1995. p. 22-27.

OLIVEIRA, J. R.P; SOARES FILHO, W. S. Situação da cultura da acerola no Brasil e ações da *Embrapa Mandioca e Fruticultura* em recursos genéticos e melhoramento. In: QUEIRÓZ, M. A; GOEDERT, C. O.; RAMOS, S. R. R. (Ed.) **Recursos genéticos e melhoramento de plantas para o nordeste brasileiro.** Petrolina: CPATSA, 1999. Anais on line. Disponível em www.cpatsa.embrapa.br.

OSBORNE, R. The use of sire and dam family averages in increasing the efficiency of selective breeding under a hierarchical mating system. **Heredity**, v. 11, p. 93-116, 1957.

PAGNACO, G. **Genetica applicata alle produzioni animali.** Milano: Città Studi Edizioni, 1997. 295 p.

PAIVA, J. R de. **Melhoramento genético de espécies agroindustriais na Amazônia: estratégias e novas abordagens.** Brasília: EMBRAPA-SPI, 1998, 135p.

PAIVA, J. R. de.; ALVES, R. E.; CORREA, M. P. F.; FREIRE, F. C. O.; BARBOSA SOBRINHO, R.; JUCA, W. Seleção e clonagem de plantas de acerola (*Malpighia* spp.). In: CONGRESSO BRASILEIRO DE FRUTICULTURA, 14. Curitiba, 1996. **Resumos.** p.38.

PAIVA, J. R. de; ALVES. R. E.; BARROS, L. M. Melhoramento genético da aceroleira (*Malpighia emarginata* DC) na Embrapa Agroindústria Tropical. In: QUEIRÓZ, M.A; GOEDERT, C.O.; RAMOS, S.R.R. (Ed.) **Recursos genéticos e melhoramento de plantas para o nordeste brasileiro.** Petrolina: CPATSA, 1999. Anais on line. Disponível em www.cpatsa.embrapa.br.

PAIVA, J. R. de; GONÇALVES, P. S. **Eficiência do programa de melhoramento da seringueira no Centro Nacional de Pesquisa de Seringueira e Dendê.** Manaus: Embrapa, 1989. (Embrapa Amazônia Ocidental, Boletim de Pesquisa , 2). 40 p.

PAIVA, J. R de.; KALIL FILHO, A. N. Melhoramento genético da seringueira. **Informe Agropecuário**, Belo Horizonte, v. 11, n. 121, p. 52-57, 1985.

PAIVA, J. R. de; PAIVA, W. O.; CORDEIRO, E. R.; SOUZA NETO, H. Parâmetros genéticos em progênies de polinização livre de acerola. **Pesquisa Agropecuária Brasileira**, v. 34, n. 4, p. 629-634, 1999.

PAIVA, J. R. de; RESENDE, M. D. V. de; CORDEIRO, E. R. Avaliação do número de colheitas na produção de progênies de acerola, repetibilidade e herdabilidade de caracteres. **Revista Brasileira de Fruticultura**, v. 23, n.1, p. 102-107, 2001.

- PAIVA, J. R. de; 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*, v. 37, n. 6, p. 799 – 807, 2002.
- PANSE, K.; SUKHATME, P. V. *Métodos estadísticos para investigadores agrícolas*. México: Fondo de Cultura Econômica, 1963. 349 p.
- PANTER, D. M.; ALLEN, F. L. Using best linear unbiased predictions to enhance breeding for yield in soybean. I. Choosing parents. *Crop Science*, v. 35, p. 379-405, 1995.
- PAPADAKIS, J. Advances in the analysis of field experiments. *Commun. Acad. Athenea*, v. 59, p. 326-342, 1984.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, London, v. 58, p. 545-554, 1971.
- PAUL, S. R.; ISLAM, A. S. Analysis of proportions in the presence of over-/under dispersion. *Biometrics*, v. 51, n. 4, p. 1400-1410, 1995.
- PEARCE, S. C. Field experimentation on rough land: the method of Papadakis reconsidered. *Journal of Agricultural Science*, v. 131, p. 1-11, 1998.
- PEARCE, S. C. *Field experimentation with fruit trees and other perennial plants*. Farnham Royal: CAB, 1976. 182p. (Technical Communication, 23).
- PEARSON, K. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans. of the Royal Society of London*: Section A, v. 200, p. 1-66, 1903.
- PERECIN, D.; BARBOSA, J. C. Uma avaliação de seis procedimentos para comparações múltiplas. *Revista de Matemática e Estatística*, v. 6, p. 95-103, 1988.
- PEREIRA, B. B. Inferência verossimilhança. *Boletim da Associação Brasileira de Estatística*, v. 38, p. 31-42, 1997.
- PEREIRA, M. B. *Comparação de métodos de seleção em populações parcialmente autógamas*. Piracicaba: ESALQ, 1989, 147 p. Tese Doutorado.
- PEREIRA, M. B.; VENCOVSKY, R. Limites da seleção recorrente. I. Fatores que afetam o acréscimo das frequências alélicas. *Pesquisa Agropecuária Brasileira*, n. 23, p. 769-780, 1988.
- PEREIRA, M.G.; PIRES, J.L.; MONTEIRO, W.R. Melhoramento do cacau. In: BORÉM, A. (Ed.). *Melhoramento de espécies cultivadas*. Viçosa: Ed. da UFV, 1999. p. 159-188.
- PESEK, J.; BAKER, R. J. Desired improvement in relation to selection indices. *Canadian Journal of Plant Science*, v. 49, p. 803-809, 1969.
- PHILLIPS, M. A.; BROWN, A. H. D. Mating system and hybridity in *Eucalyptus pauciflora*. *Australian Journal of Biological Science*, v. 30, p. 337-344, 1977.

PIMENTEL GOMES, F. Novos aspectos do problema do tamanho ótimo de parcelas em experimentos com plantas arbóreas. *Pesquisa Agropecuária Brasileira*, v. 23, n. 1, p. 59-62, 1988.

PIMENTEL GOMES, F. *Curso de estatística experimental*. 12. ed. São Paulo: Nobel, 1987.

PIMENTEL GOMES, F. O problema do tamanho das parcelas em experimentos com plantas arbóreas. *Pesquisa Agropecuária Brasileira*, v. 19, n. 12, p. 1507-1512, 1984.

PIMENTEL GOMES, F.; CHAVES, R. A amostragem ótima em inventário florestal. *IPEF*, v. 38, p. 17-22, 1988.

PIMENTEL GOMES, F.; COUTO, H. T. Z. O tamanho ótimo de parcela experimental para ensaios com eucaliptos. *IPEF*, Piracicaba, v. 31, p. 75-77, 1985.

PINTO JÚNIOR, J. E. *Variabilidade genética em populações de Eucalyptus urophylla*. Piracicaba: ESALQ, 1984. Tese Mestrado.

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.

PLACKETT, R. L. A class of bivariate distributions. *Journal of the American Statistical Association*, v. 60, p. 516-522, 1965.

POLLAK, E. Effective population numbers and mean times to extinction in dioecious populations with overlapping generations. *Mathematical Bioscience*, v. 52, p. 1-25. 1980.

PONZONI, R. W. Objectives and selection criteria for Australian Merino sheep. *Proceedings of the Australian Association of Animal Breeding and Genetics*, v. 1, p. 320-336, 1979.

PONZONI, R. W. *Objetivos y criterios de selección en lanares*. Secretariado Uruguai de la Lana. Montevideo: Cone Sur, 1986, p. 79-94. (Ovinos e Lanas, Boletim Técnico, 15).

PORTNOY, S. Maximizing the probability of correctly ordering random variables using linear predictors. *Journal of Multivariate Analysis*, v. 12, p. 256-269, 1982.

POWELL, M. B.; NIKLES, D. G. Genetic parameter estimates and predicted breeding values for diameter, height and stem straightness of *Pinus elliotii*, *Pinus caribaea* var *hondurensis* and their F1 hybrids. In: DIETERS, M. J.; MATHESON, A. C.; NIKLES, D. G.; HARWOOD, C.; WALKER, S. M. (Ed.). *Tree improvement for sustainable tropical forestry*. Caloundra: QFRI / IUFRO, 1996. p. 169-172.

PREECE, D. A. The design and analysis of experiments: what has gone wrong? *Utilitas Mathematica*, v. 21, A, p. 201-244, 1982.

- PURI, S.; RAJORA, N. Influence of physiology and environment in raising trees in mass by vegetative propagation techniques. In: LAMBETH, C.; DVORAK, W. **Breeding tropical trees**. IUFRO CONFERENCE, 1992. Cali. Proceedings. Raleigh: North Carolina State University, 1992. p. 311.
- QIAO, C. G.; BASFORD, K. E.; DELACY, I. H.; COOPER, M. Evaluation of experimental designs and spacial analysis in wheat breeding trials. **Theoretical and Applied Genetics**, v. 100, p. 9-16, 2000.
- QUAAS, R. L. Additive genetic model with groups and relationships. **Journal of Dairy Science**, v. 71, p. 1338-1345, 1988.
- QUAAS, R. L.; POLLAK, E. J. Mixed model methodology for farm and ranch beef cattle testing programs. **Journal of Animal Science**, v. 51, p. 1277-1287, 1980.
- QUAAS, R. L.; POLLAK, E. J. Modified equations for sire models with groups. **Journal of Dairy Science**, v. 64, p. 1868-1872, 1981.
- QUENOUILLE, M. H. Notes on bias in estimation. **Biometrika**, v. 43, p. 353-360, 1956.
- QUINTON, M.; SMITH, C.; GODDARD, M. E. Comparison of selection methods at the same level of inbreeding. **Journal of Animal Science**, v. 70, p. 1060-1067, 1992.
- RADEN, P. M. van. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. **Journal of Dairy Science**, v. 75, p. 3136-3144, 1992.
- RADEN, P. M. van; HOESCHELE, I. Rapid inverse of additive by additive relationship matrices by including sire-dam combination effects. **Journal of Dairy Science**, v. 74, p. 570-579, 1991.
- RADEN, P.M. van; WIGGANS, G. R. Derivation, calculation and use of national animal model information. **Journal of Dairy Science**, v. 74, p. 2737-2746, 1991.
- RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: Editora UFLA, 2000. 303 p.
- RAMALHO, M. A. P.; GONÇALVES, F. M. A.; SOUZA SOBRINHO, F. Emprego da seleção recorrente no melhoramento do cafeeiro. In: NUMBERG, P.L. et al. SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 3., 1999, Lavras. **Anais...** Lavras: Editora UFLA, 1999. p. 66-81.
- RAMALHO, M. A. P.; SANTOS, J. B.; ZIMMERMANN, M. J. O. **Genética quantitativa no melhoramento de plantas autógamas**. Goiânia: Universidade Federal de Goiás, 1993. 271p.
- RAMALHO, M. A. P.; VENCOVSKY, R. Estimação dos componentes de variância genética em plantas autógamas. **Ciência e Prática**, v. 2, n. 2, p. 117-140, 1978.

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

RAWLINGS, J. O. Present status of research on long and short-term recurrent selection in finite populations: choice of population size. In: MEETING OF WORKING GROUP ON QUANTITATIVE GENETICS, 2., 1970, New Orleans. **Proceedings...** USDA / SFES, 1970. p. 1-15.

REDDY, K. V.; ROCKWOOD, D. L. Breeding strategies for coppice production in a Eucalyptus grandis base population with four generation of selection. **Silvae Genetica**, v. 38, p. 148-151, 1989.

REKAYA, R. **Analisis Bayesiano de datos de producción en los días de control para la selección de caracteres lecheros**. 1997. Madrid: Universidad Politécnica de Madrid. 183p. Doctoral Thesis.

RESENDE, M. D. V. de. Avanços da genética biométrica florestal. In: BANDEL, G.; VELLO, N. A.; MIRANDA FILHO, J. B. (Ed.). **Encontro sobre temas de genética e melhoramento**: genética biometria vegetal. **Anais**. Piracicaba: Esalq/Usp, 1997. p.20-46.

RESENDE, M. D. V. de. Genética biométrica no melhoramento de fruteiras. In: PAIVA, J. R. et al. (Ed.). **Encontro de Genética do Nordeste**. Fortaleza: Embrapa Agroindústria Tropical, 2000. Apresentado ENGENE, 15., 2000, Fortaleza. Anais on line. Disponível em www.cnpat.embrapa.br.

RESENDE, M. D. V. de. Correções nas expressões do progresso genético com seleção em função da amostragem finita dentro de famílias e populações e implicações no melhoramento florestal. **Boletim de Pesquisa Florestal**, Colombo, v. 22/23, p. 61-77, jan./dez. 1991.

RESENDE, M. D. V. de. Correlação intraclasse entre valores genéticos e implicações na comparação de métodos de seleção. **Floresta**, Curitiba, v. 24, n. 1/2, p.37-48, jun./dez. 1994.

RESENDE, M. D. V. de. Delineamento de experimentos de seleção para a maximização da acurácia seletiva e progresso genético. **Revista Árvore**, v. 19, n. 4, p. 479-500, out./dez. 1995.

RESENDE, M. D. V. de. **Efeitos fixos ou aleatórios de repetições no contexto dos modelos mistos no melhoramento de plantas perenes**. Colombo: Embrapa Florestas, 2002. 15 p. (Embrapa Florestas. Documentos, 68).

RESENDE, M. D. V. de. Interação genótipo x ambiente e determinação do número adequado de locais de experimentação com base nas estatísticas *F* de Snedecor da análise de variância conjunta. **Boletim de Pesquisa Florestal**, Colombo, v. 37, p. 55-66, jul./dez. 1998.

RESENDE, M. D. V. de. Melhoramento de essências florestais. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: UFV, 1999a. p. 589-648.

RESENDE, M. D. V. de. Numerical linear algebra problems arising from computational forest genetics. In: INTERNATIONAL WORKSHOP ON NUMERICAL LINEAR ALGEBRA AND OPTIMIZATION, 1., 1997. Curitiba. *Proceedings...* Curitiba: UFPR, 1997a. p. 58-60.

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, 1999b. 434 p. Tese Doutorado.

RESENDE, M. D. V. de. *Seleção de genótipos de milho (*Zea mays L.*) em solos contrastantes*. Piracicaba: USP/ESALQ, 1989. 230 p. Dissertação Mestrado.

RESENDE, M. D. V. de. Seleção precoce no melhoramento florestal. In: WORKSHOP SIF/UNIVERSIDADE FEDERAL DE VIÇOSA: MÉTODOS DE SELEÇÃO, 1, 1994, Belo Horizonte. *Anais...* Viçosa: Sociedade de Investigações Florestais, 1994. p. 58-73.

RESENDE, M. D. V. de; ARAÚJO, A. J. Modelo genético-estatístico para estimação de componentes de variação genética e parâmetros genéticos em testes de progênies com indivíduos repetidos clonalmente. *Floresta*, Curitiba, v. 23, n. 1/2, p. 35-46, 1993.

RESENDE, M. D. V. de.; ARAÚJO, A. J.; SAMPAIO, P. T. B.; WIECHETECK, M. S. S. Acurácia seletiva, intervalos de confiança e variâncias de ganhos genéticos associados a 22 métodos de seleção em *Pinus caribaea* var. *hondurensis*. *Floresta*, Curitiba, v. 25, n. 1/2, p. 3-16, 1995b.

RESENDE, M. D. V. de; BERTOLUCCI, F. L. G. Maximization of genetic gain with restriction on effective population size and inbreeding in *Eucalyptus grandis*. In: IUFRO CONFERENCE "EUCALYPT PLANTATIONS: Improving Fibre Yield And Quality", 1995, Hobart. *Proceedings...* Hobart: IUFRO-CRCTHF, 1995. p. 167-170.

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

RESENDE, M. D. V. de; DUDA, L. L. ; GUIMARÃES, P. R. B.; FERNANDES, J. S. C. Análise de modelos lineares mistos via inferência Bayesiana. *Revista de Matemática e Estatística*. São Paulo, v. 19, p. 41-70, 2001b.

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; FERNANDES, J. S. C. Análises alternativas envolvendo o procedimento BLUP e o delineamento experimental de blocos incompletos ou látice. **Revista de Matemática e Estatística**. São Paulo, v. 18, p.103-124, 2000.

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; FREITAS, A. R. Estimadores para a repetibilidade envolvendo efeito materno em caracteres de bovinos de corte. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 1998, Botucatu. **Anais...** Botucatu: UNESP, 1998. v. 3, p. 234-236.

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**, Campinas, v. 60, n.3, p. 185 - 193, 2001c.

RESENDE, M. D. V. de; HIGA, A. R. Aplicación de técnicas de análisis multivariado en el estudio de la divergencia genética entre procedências de *E. viminalis*. In: JORNADAS SOBRE EUCALIPTOS DE ALTA PRODUCTIVIDAD, 1991, Buenos Aires. **Actas**. Buenos Aires: Centro de Investigaciones y Experiencias Forestales, 1991. v. 1, p. 139-154.

RESENDE, M. D. V. de; HIGA, A. R. **Cooperación técnica em melhoramento genético de Eucalyptus grandis**. Colombo: EMBRAPA / CENIBRA, 1992. 8 p. Relatório técnico não publicado.

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. Estratégias de melhoramento para eucalipto visando a seleção de híbridos. **Boletim de Pesquisa Florestal**, Colombo, v. 20/21, p. 1-20, 1990.

RESENDE, M. D. V. de; HIGA, A. R. Maximização da eficiência da seleção em testes de progêniens de *Eucalyptus* através da utilização de todos os efeitos do modelo matemático. **Boletim de Pesquisa Florestal**, Colombo, v. 28/29, p. 37-55, 1994a.

RESENDE, M. D. V. de; HIGA, A. R.; AUER, C. G.; HELLER, J. B. Análise binomial da resistência genética a gomose em acacia-negra. **Revista Brasileira de Genética** (Suplemento), Ribeirão Preto, v. 16, n. 3, p. 369, 1993a.

RESENDE, M. D. V. de; HIGA, A.R.; HELLER, J.B; STEIN, P.P. Parâmetros genéticos e interação genótipo x ambiente em teste de procedência e progêniens de acácia-negra (*Acacia mearnsii De Wild*). **Boletim Pesquisa Florestal**, Colombo, 24/25. p. 55-65, 1992.

- 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...*, Curitiba: SBS, 1993. p. 144-147.
- RESENDE, M. D. V. de; HIGA, A. R.; LAVORANTI, O. J. Regressão geno-fenotípica multivariada e maximização do progresso genético em programas de melhoramento de *Eucalyptus*. *Boletim de Pesquisa Florestal*, Colombo, n. 28/29, p. 57-71, 1994b.
- RESENDE, M. D. V. de; MEDRADO, M. J. S. Aspectos metodológicos no melhoramento genético de *Leucaena leucocephala*, uma espécie florestal autógamma. In: ENCONTRO SOBRE SISTEMAS AGROFLORESTAIS NOS PAISES DO MERCOSUL, 1., 1994, Porto Velho. *Anais...* Colombo, 1994, v. 2, p. 233-248. (EMBRAPA-CNPF. Documentos, 27).
- 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*, Curitiba , v. 28 , n. 1/2 , p. 51-63, 1998a.
- RESENDE, M. D. V. de; MORAES, M. L. T.; PAIVA, J. R. Comparação de métodos de avaliação genética e seleção no melhoramento da seringueira. *Floresta*, Curitiba, v. 26, n. 1/2, p. 25-47, 1996d.
- RESENDE, M. D. V. de; OLIVEIRA, E. B. DE; MELINSKI, L. C.; GOULART, F. S.; OAIDA, G. R. SELEGEN - *Seleção Genética Computadorizada*: manual do usuário. Colombo: EMBRAPA-CNPF, 1994a, 31 p.
- RESENDE, M. D. V. de; OLIVEIRA, E. B. Sistema “SELEGEN” – Seleção Genética Computadorizada para o melhoramento de espécies perenes. *Pesquisa Agropecuária Brasileira*, Brasília, v. 32, n. 9, p. 931-939, 1997.
- RESENDE, M. D. V. de; OLIVEIRA, E. B.; HIGA, A. R. Utilização de índices de seleção no melhoramento do *Eucalyptus*. *Boletim de Pesquisa Florestal*, Colombo, n. 21, p. 1-13, 1990.
- 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*. *Boletim de Pesquisa Florestal*, Colombo, n. 32/33, p. 18-45, jan./dez. 1996b.
- RESENDE, M. D. V. de; PRATES, D. F.; JESUS, A.; YAMADA, C. K. Melhor predição linear não viciada (BLUP) de valores genéticos no melhoramento de *Pinus*. *Boletim de Pesquisa Florestal*, n. 32/33, p. 3-22, jan./dez. 1996a.
- RESENDE, M. D. V. de; RESENDE, R. M. S. Seleção de cruzamentos, recombinação desbalanceada e aumento da eficiência do melhoramento genético de *Eucalyptus*. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCLYPT, 1997, Salvador. *Proceedings...* Colombo: Embrapa Florestas, 1997. v.1, p. 377-341.

RESENDE, M. D. V. de; REZENDE, G. D. S. P.; FERNANDES, J. S. C. Regressão aleatória e funções de covariância na análise de medidas repetidas. **Revista de Matemática e Estatística**, São Paulo, v. 19, p. 21-40, 2001a.

RESENDE, M. D. V. de; ROSA-PEREZ, J. R. H. **Genética quantitativa e estatística no melhoramento animal**. Curitiba: Imprensa Universitária - UFPR, 1999a. 496 p.

RESENDE, M. D. V. de; ROSA-PEREZ, J. R. H. Melhoramento animal: predição de valores genéticos pelo modelo animal (BLUP) em bovinos de leite, bovinos de corte, ovinos e suínos. **Archives of Veterinary Science**, v. 4, n. 1, p. 17-30, 1999b.

RESENDE, M. D. V. de; SILVA, H. D. Estratégia de melhoramento para a erva-mate baseada no coeficiente de repetibilidade. In: CONGRESSO FLORESTAL E DO MEIO-AMBIENTE DO PARANÁ, 3., 1991, Curitiba. **Anais...** Curitiba: Associação Paranaense de Engenheiros Florestais, 1991. p. 241-251.

RESENDE, M. D. V. de; SIMEÃO, R. M.; FERNANDES, J. S. C. BLUP individual com medidas repetidas aplicado ao melhoramento de espécies perenes. **Ciência e Agrotecnologia**, v. 24, n. 4, p. 986-997, 2000a.

RESENDE, M. D. V. de; SIMEÃO, R. M.; FERNANDES, J. S. C.; STURION, J. A. Melhoramento e seleção em erva-mate (*Ilex paraguariensis*): contribuição e experiências de um século de melhoramento do chá-da-índia (*Camellia sinensis*). **Boletim de Pesquisa Florestal**, v. 37, p. 67-80, 1998b.

RESENDE, M. D. V. de; SIMEÃO, R. M.; STURION, J. A. Fundamentos de genética de populações para conservação de germoplasma de erva-mate. In: CONGRESSO SUL AMERICANO DA ERVA-MATE, 1, 1997, Curitiba. **Anais**. Colombo: EMBRAPA, 1997b. v.1. p.173-207.

RESENDE, M. D. V. de; STURION, J. A. **Análise genética de dados com dependência espacial e temporal no melhoramento de plantas perenes via modelos geoestatísticos e de séries temporais empregando REML/BLUP individual**. Colombo: Embrapa Florestas, 2001. 80 p. (Embrapa Florestas. Documentos, 65).

RESENDE, M. D. V. de; STURION, J. A.; CARVALHO, A. P.; SIMEÃO, M. R.; FERNANDES, J. S. C. **Programa de Melhoramento da Erva-mate Coordenado pela Embrapa**: resultados da avaliação genética de populações, progêniens, indivíduos e clones. Colombo: Embrapa Florestas, 2000b. 65 p. (Embrapa Florestas. Circular Técnica, 43).

RESENDE, M. D. V. de; STURION, J. A.; MENDES, S. **Genética e melhoramento da erva-mate (*Ilex paraguariensis St. Hill*)**. Colombo: EMBRAPA-CNPQ, 1995a. 33p. (EMBRAPA-CNPQ. Documentos, 25).

RESENDE, M. D. V. de; STURION, J. A.; PRATES, D. F.; MENDES, S. Tamanho de parcela, estimativa de parâmetros e ganhos genéticos em erva-mate sem a utilização de testes de progêniens. **Boletim de Pesquisa Florestal**, Colombo, v. 32/33 , p. 53 -66 , 1996c.

- RESENDE, M. D. V. de; STURION, J. A.; SIMEÃO, R.M. Estratégias para o melhoramento genético de erva-mate. In: CONGRESSO SUL-AMERICANO DA ERVA-MATE, 1.; REUNIAO TÉCNICA DO CONE SUL SOBRE A CULTURA DA ERVA-MATE, 2., 1997, Curitiba. *Anais..* Colombo: EMBRAPA-CNPF, 1997a. p. 243-266.
- RESENDE, M. D. V. de; VENCOVSKY, R. Condução e utilização de bancos de conservação genética de espécies de eucalipto. In: CONGRESSO FLORESTAL BRASILEIRO, 6., 1990, Campos do Jordão. *Anais...* São Paulo: Sociedade Brasileira de Silvicultura, 1990. p. 434-439.
- RESENDE, M. D. V. de; VENCOVSKY, R. Numerical evaluation of the relative efficiency of *Eucalyptus* selection methods based on half-sib and selfed families. I. Progeny Testing. In: IUFRO CONFERENCE-BREEDING TROPICAL TREES, Cali, 1992. *Proceedings...* Raleigh: North Carolina State University, 1992. p. 175-190.
- RESENDE, M. D. V. de; VENCOVSKY, R. Relative efficiency of *Eucalyptus* selection methods. In: INTERNATIONAL CONGRESS OF GENETICS, 7., 1993, Birmingham. *Abstracts.* Birmingham: University of Birmingham, 1993. p. 226.
- 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, 1995c. p. 191-193.
- RESENDE, M. D. V. de; YUAN, J. Y.; JESUS, A.; YAMADA, C.K. Método interativo de resolução de sistemas de equações lineares aplicado à predição de valores genéticos em espécies perenes. REUNIÃO DA ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA, 28, Curitiba, 1996. *Resumos...* Curitiba: Universidade Federal do Paraná, 1996. p. 29.
- REVERTER, A. Empirical evidence of the optimality of method R estimates. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. *Proceedings...* Armidale: AGBU/University of New England, 1998. v. 25, p. 533-536.,
- REVERTER, A.; GOLDEN, B. L.; BOURDON, R. M.; BRINKS, J. S. Method R Variance Components Procedure: Application on the simple breeding value model. *Journal of Animal Science*, v. 72, p. 2247-2253, 1994.
- REZENDE, G. D. P. S.; BERTOLUCCI, F. L. G. Uso da seleção combinada na determinação da eficiência da seleção precoce em progêniens de meios irmãos de *Eucalyptus urophylla*. In: CONGRESSO FLORESTAL BRASILEIRO, 7., 1993, Curitiba. *Anais.* Curitiba: SBS, 1993. p. 158-160.
- REZENDE, G. D. S. P.; RESENDE, M. D. V. de. Dominance effects in *Eucalyptus grandis*, *Eucalyptus urophylla* and hybrids. In: DUNGEY, H. S.; DIETERS, M. J.; NIKLES, D. G. (Eds.). Hybrid Breeding and Genetics of Forest Trees, 2000, Noosa. *Proceedings...* Brisbane: Department of Primary Industries, 2000. p. 93-100.

RIBEIRO JUNIOR, P. J. *Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo*. 1995. 99 f. Dissertação (Mestrado) – ESALQ/USP, Piracicaba.

RITLAND, K. A. Marker-based method for inferences about quantitative inheritance in natural populations. *Evolution*, v. 50, p. 1062-1073, 1996.

RITLAND, K.; JAIN, S. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. *Heredity*, v. 47, n. 1, p. 35-52, 1981.

ROBERTSON, A. A numerical description of breed structure. *Journal of Agricultural Sciences*, v. 43, p. 334-336, 1953.

ROBERTSON, A. A theory of limits in artificial selection with many linked loci. In: KOJIMA, K. (Ed.). *Mathematical topics in population genetics*. Berlin: Springer, 1970. p. 246-288.

ROBERTSON, A. A theory of limits in artificial selection. *Proceedings of the Royal Society of London B.*, London, v. 153, p. 234-249, 1960.

ROBERTSON, A. Artificial selection with a large number of linked loci. In: POLLAK, E.; KEMPTHORNE, A.; BAILEY, T. B. (Ed.). INTERNATIONAL CONFERENCE ON QUANTITATIVE GENETICS, 1977, Ames. *Proceedings* ... Ames: Iowa State University Press, 1977. p. 307-322.

ROBERTSON, A. Experimental design in the evaluation of genetic parameters. *Biometrics*, Washington, v. 15, p. 219-226, 1959.

ROBERTSON, A. Inbreeding in artificial selection programmes. *Genetical Research*, Cambridge, v. 2, p. 189-194, 1961.

ROBERTSON, A. Prediction equations in quantitative genetics. *Biometrics*, Washington, v. 11, p. 95-98, 1955.

ROBERTSON, A. Weighting in estimation of variance components. *Biometrics*, Washington, v. 18, p. 413-415, 1962.

ROBERTSON, A.; ASKER, A. A. The genetic history and breed-structure of British Friesian cattle. *Empire Journal of Experimental Agriculture*, v.19, p.113-130, 1951.

ROBERTSON, A.; HILL, W. G. Population and quantitative genetics of many linked loci in finite populations. *Proceedings of the Royal Society of London B.*, London, v. 219, p. 253-264, 1983.

ROBERTSON, A.; LERNER, I. M. The heritability of all-or-none traits: viability of poultry. *Genetics*, Tokyo, v. 34, p. 395-411, 1949.

ROBINSON, D. L. Estimation and use of variance components. *The Statistician*, Londres, v. 36, p. 3-15, 1987.

ROBINSON, D. L. *Estimation of variance components in non-orthogonal data by residual maximum likelihood*: REML User Manual. Edinburgh: Scottish Agricultural Statistics Service. 1987. 33 p.

- ROBINSON, D. L. That BLUP is a good thing: the estimation of random effects. *Statistical Science*, Hayward, v. 6, p. 15-32, 1991.
- RODRIGUES-ZAS, S. L.; GIANOLA, D.; SHOOK, G. E. Factors affecting susceptibility to intramammary infection and mastitis: an approximate Bayesian analysis. *Journal of Dairy Science*, Champaign, v. 80, p. 75-85, 1997.
- RODRIGUEZ, M. C.; TORO, M.; SILIO, L. Selection on lean growth in a nucleus of Landrace pigs: an analysis using Gibbs sampling. *Animal Science*, Penicuik, v. 63, p. 243-253, 1996.
- RODRIGUEZ-ALMEIDA, F. A.; VAN VLECK, L. D.; WILLHAM, R. L.; NORTHCUTT, S. L. Estimation of non-additive genetic variances in three synthetic lines of beef cattle using an animal model. *Journal of Animal Science*, Champaign v. 73, p. 1002-1011, 1995.
- ROGERS, W. H.; TUKEY, J. W. Understanding some long-tailed distributions. *Statistica Neerlandica*, v. 26, p. 211-226, 1972.
- ROMAN, R. M.; WILCOX, C. J.; MARTIN, F. G. Estimates of repeatability and heritability of productive and reproductive traits in a herd of Jersey cattle. *Genetics and Molecular Biology*, Ribeirão Preto, v. 23, n. 1, p. 113-120, 2000.
- RONNINGEN, K. Some properties of the selection index derived by "Henderson's mixed model method". *Z. Tierz Zuchungsbiol*, v. 88, p. 186, 1971.
- RONNINGEN, K.; VAN VLECK, L. D. Selection index theory with practical applications. In: CHAPMAN, A. B. (Ed.). *World Animal Science A4: General and quantitative genetics*. Oxford: Elsevier Science Publ., 1985, p. 187-225.
- ROSA, G. J. M. Robust mixed linear models in quantitative genetics: Bayesian analysis via Gibbs sampling. In: INTERNATIONAL SYMPOSIUM ON ANIMAL BREEDING AND GENETICS. 1999. Viçosa. *Proceedings...* Viçosa: Ed. da Universidade Federal de Viçosa, 1999. p. 133-159.
- ROSA-PEREZ, J. R. H.; RESENDE, M. D. V. de. Importância da genética e do melhoramento para a produção animal. In: SIMPÓSIO SOBRE GENÉTICA E MELHORAMENTO ANIMAL, 1, 1999, Palotina. *Anais...* Curitiba: Ed. da Universidade Federal do Paraná, 1999. p. 1-21.
- SAIO, K.; HAYASHI, T. Bayesian estimation of genetic parameters. *Japanese Journal of Breeding*, v. 40, p. 63-75, 1990.
- SAKIYAMA, N. S. Melhoramento do café arábica. In: BORÉM, A. (Ed.). *Melhoramento de espécies cultivadas*. Viçosa: Imprensa Universitária, 1999. p. 189-204.
- SALES, J.; HILL, W. G. Effect of sampling errors on efficiency of selection indices. 1. Use of information from relatives for single trait improvement. *Animal Production*, v. 22, p. 1-17, 1976a.

SALES, J.; HILL, W. G. Effect of sampling errors on efficiency of selection indices. 2 Use of information on associated traits for improvement of a single trait. *Animal Production*, v. 23, p. 1-14, 1976b.

SAMPAIO, P. T. B. *Variação genética entre procedências e progêneres de Pinus oocarpa, Pinus caribaea var. hondurensis e Pinus maximinoi e métodos de seleção para melhoramento genético*. Curitiba: UFPR, 1996. 169 p. Tese Doutorado.

SAMPAIO, P. T.; RESENDE, M. D. V.; ARAÚJO, A. J. Estimativas de parâmetros genéticos e métodos de seleção para o melhoramento genético de *Pinus caribaea* var. *hondurensis*. *Pesquisa Agropecuária Brasileira*, Brasília, v. 35 , n. 11, p. 2243 – 2253, 2000.

SANTOS, P. E. T.; GARCIA, C. H.; MORI, E. S.; MORAES, M. L.T. Potencial para programas de melhoramento, estimativas de parâmetros genéticos e interação progêneres x locais em populações de *Eucalyptus urophylla*. *IPEF*, Piracicaba, v. 33/44, p. 11-19, 1990.

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometric Bulletin*, v. 2, n. 1, p. 10-14, 1946.

SAVOLAINEN, O. Genetic variation and fitness: conservation lessons from pines. In: LOESCHCKE, V.; TOMIUK, J.; JAIN, S. K. (Ed.) *Conservation genetics*. Basel: Birkhäuser Verlag, 1994. p. 27-36

SCHAEFFER, L. R. Maximum likelihood estimation of variance components in dairy cattle breeding research. *Journal of Dairy Science*, Champaign, v. 59, p. 2146-2151, 1976.

SCHAEFFER, L. R.; DEKKERS, J. C. M. Random regressions in animal models for test-day production in dairy cattle. WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 5., 1994, Guelph. *Proceedings...* Guelph: University of Guelph, 1994, v.18, p.443.

SCHAEFFER, L. R.; MAO, I. L. Notes on linear model theory and variance component estimation. In: CURSO DE GENÉTICA ANIMAL, 1988, Zaragoza. **Course notes**. Zaragoza: Universidade de Zaragoza, Centro Internacional de Altos Estudios Agronomicos Mediterrâneos. Instituto Agronômico Mediterrâneo de Zaragoza, 1988. 237 p.

SCHAEFFER, L. R.; WILTON, J. W.; THOMPSON, R. Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics*, v. 34, p. 199-208, 1978.

SCHALL, R. Estimation in generalized linear models with random effects. *Biometrika*, London, v. 78, p. 719-727, 1991.

SCHEFFÉ, H. *The analysis of variance*. New York: Wiley, 1959. 477 p.

- SCHENKEL, F. S. Effects of selection on animal genetic evaluation. In: INTERNATIONAL SYMPOSIUM ON ANIMAL BREEDING AND GENETICS. 1999. Viçosa. *Proceedings...* Viçosa: Ed. da Universidade Federal de Viçosa, 1999. p. 79-100.
- SCHNEEBERGER, M.; BARWICK, S. A.; CROW, G. H.; HAMMOND, K. Economic indices using breeding values predicted by BLUP. *Journal of Animal Breeding and Genetics*, v. 109, p. 180-187, 1992.
- SCHOENBERG, I.J. Contributions to the problem of approximation of equidistant data by analitic functions. *Quarterly Applied Mathematics*, v. 4, p. 44-99, 1946.
- SCHUTZ, W. M.; COCKERHAM, C. C. The effect of field blocking on gain from selection. *Biometrics*, v. 22, p. 843-862, 1966.
- SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, v. 30, n. 3, p. 507-512, 1974.
- SEARLE, S. R. C.R. Henderson, the statistician: his contributions to variance components estimation. *Journal of Dairy Science*, v. 74, p. 4035-4044, 1991.
- SEARLE, S. R. The value of indirect selection: I. Mass selection. *Biometrics*, v. 21, p. 682-707, 1965.
- SEARLE, S. R. Topics in variance component estimation. *Biometrics*, Washington, v. 27, p. 1-76, 1971.
- SEARLE, S. R.; CASELLA, G.; Mc CULLOCH, C. E. *Variance components*. New York: J. Wiley. 1992. 528 p.
- SERA, T. Desafios no melhoramento genético do cafeeiro. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 2., 1998, Lavras. *Anais...* Lavras: UFLA, 1998. p.105-122.
- SHAH, A.; LAIRD, N.; SCHOENFELD, D. A random effects model for multiple characteristics with possible missing data. *Journal of the American Statistical Association*, v. 92, p. 775-779, 1997.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, p. 591-611, 1965.
- SHAW, R. G. Maximum-likelihood approaches to quantitative genetics of natural populations. *Evolution*, Lawrance, v. 41, p. 812-826, 1987.
- SHELBOURNE, C. J. A. Interspecific hybrid breeding schemes. In: MEETING RESEARCH WORKING GROUP ON FOREST GENETICS, 12, 1993, Canberra. *Proceedings...* Canberra: ANU-ACT, 1993, p. 92-93.
- SHELBOURNE, C. J. A. Genetic gains from diffrent kinds of breeding populations and seed or plant production population. *South African Forestry Journal*, n. 160, p. 49-65, 1992.

- SHEPHERD, R. K. Three-tier open nucleus breeding schemes. *Animal Science*, v. 65, p. 321-324, 1997.
- SHEPHERD, R. K.; KINGHORN, B. P. Optimizing multi-tier open nucleus breeding schemes. *Theoretical and Applied Genetics*, v. 85, p. 372-378, 1992.
- SHERIDAN, A. K. Agreement between estimated and realised genetic parameters. *Animal Breeding Abstracts*, v. 56, p. 877-889, 1988.
- SHIMIZU, J. Y.; KAGEYAMA, P. Y.; HIGA, A. R. *Procedimentos e recomendações para estudos de progêniens de essências florestais..* Colombo: EMBRAPA-URPFCS, 1982. 34 p. (EMBRAP-URPFCS. Documentos, 11).
- SHRIKHANDE, V. J. Some considerations in designing experiments on coconut trees. *Journal of the Indian Society of Agricultural Statistics*, v. 9, p. 82-99, 1957.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. Tokyo: McGraw-Hill, 1956. 312 p.
- SILVA, E. C.; FERREIRA, D. F.; BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo. *Ciência e Agrotecnologia*, v. 23, n. 3, p. 687-696, 1999.
- SILVEY, S. D. *Statistical inference*. 2. ed. London: Chapman & Hall, 1975. 191 p.
- SIMMONDS, N. W. *Principles of crop improvement*. New York: Longman, 1979. 386 p.
- SIMMONDS, N. W. Rubber breeding. In: WEBSTER, C.C.; BAULKWILL, W.J. (Ed.). *Rubber*. Longman: New York, 1989. p.85-124.
- SIMONS, A. J. Genetic improvement of non-industrial trees. *Agroforestry Systems*, v. 18, p. 197-212, 1992.
- SIQUEIRA, E. R. Coeficiente de repetibilidade da produção de frutos de coqueiro comum. *Pesquisa Agropecuária Brasileira*, v. 17, n. 3, p. 573-574, 1982.
- SIQUEIRA, E. R.; RIBEIRO, F. E.; ARAGÃO, W. M.; TUPINAMBÁ, E. A. Melhoramento genético do coqueiro. In: FERREIRA, J.M.S.; WARWICK, D.R.N.; SIQUEIRA, L.A. *A cultura do coqueiro no Brasil*. Brasília: Embrapa-SPI, 1998. p. 73-98.
- SMITH, A. F. M. Bayesian statistics. Present position and potential developments: some personal views. *Journal of the Royal Statistical Society: Series A*, v. 147, p. 245-259, 1984.
- SMITH, C. Improvement of metric traits through specific loci. *Animal Production*, v. 9, p. 349-358, 1967.

- SMITH, C. Potential for animal breeding, current and future. In: WEIR, B. S.; EISEN, E. J.; GOODMAN, M. M.; NAMKOONG, G. (Ed.). International Conference on Quantitative Genetics, 2., Raleigh, 1988. *Proceedings...* Sunderland: Sinauer, 1988. Chapter 13.
- SMITH, C. The use of specialized sire and dam lines in selection for meat production. *Animal Production*, v. 6, p. 337-344, 1964.
- SMITH, C.; QUINTON, M. The effect of selection in sublines and crossing on genetic response and inbreeding. *Journal of Animal Science*, v. 71, p. 2.631-2.638, 1993.
- SMITH, F. H. An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, n. 28, p. 1-23, 1938.
- SMITH, O. S. Discussion Session. In: COORS, J.G.; PANDEY, S. (Ed.). *The genetics and exploitation of heterosis in crops*. Madison: ASA / CSSA / SSSA, 1999. p. 283.
- SMITH, S. P.; GRASER, H. U. Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science*, Champaign, v. 69, p. 1156-1165, 1986.
- SNEDDEN, C. L.; VERRYN, S.; ROUX, C. Z. Broad and narrow sense heritabilities in a cloned open pollination *Eucalyptus grandis* breeding population. In: IUFRO CONFERENCE ON FOREST GENETICS FOR THE NEXT MILLENIUM, Durban, 2000. *Proceedings...* Durban: IUFRO-ICFR, 2000. p. 214-220.
- SNEDECOR, G. W.; COCHRAN, W. G. *Statistical methods*. 6. ed. Iowa: Iowa State University Press, 1967. 507p.
- SOARES, A. *Geoestatística para as ciências da terra e do meio ambiente*. Lisboa: IST Press, 2000. 206 p.
- SOH, A. C. Ranking parents by best linear unbiased prediction (BLUP) of breeding values in oil palm. *Euphytica*, v. 76, n. 1-2, p. 13-21, 1994.
- SOHN, S.; GODDARD, R. E. Influence of infection percent on improvement of fusiform rust resistance in slash pine. *Silvae Genetica*, v. 28, p. 5-6, 1979.
- SORENSEN, D. A. *Gibbs sampling in quantitative genetics*. Tjele: Danish Institute of Animal Science, Department of Breeding and Genetics. Intern Report n.82. 1996. 186 p.
- SORENSEN, D. A.; KENNEDY, B. W. The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. *Theoretical and Applied Genetics*, v. 66, p. 217-220, 1983.
- SORENSEN, D. A.; WANG, C. S.; JENSEN, J.; GIANOLA, D. Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genetics Selection Evolution*, v. 26, p. 333-360, 1994.

SORENSSON, C. T. Planting designs for profit-oriented Leucaena hybrid seed production. In: LAMBETH, C.; DVORAK, W. **Breeding Tropical Trees**. IUFRO CONFERENCE, 1992, Cali. **Proceedings**. Raleigh: North Carolina State University, 1992. p.359-365.

SORIA, F.; BASURCO, F.; TOVAL, G.; SILIO, L.; RODRIGUEZ, M. C.; TORO, M. An application of bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. **Canadian Journal of Forestry Research**, v. 28, p. 1286-1204, 1998.

SOUZA JUNIOR, C. L. Comparisons of intra, interpopulation and modified recurrent selection methods. **Brazilian Journal of Genetics**, v. 16, n. 1, p. 91-105, 1993.

SOUZA JUNIOR, C. L. **Melhoramento de espécies de reprodução vegetativa**. Piracicaba: ESALQ, Departamento de Genética, 1995. 41 p.

SOUZA JUNIOR, C. L. Recurrent selection and heterosis. In: COORS, J.G.; PANDEY, S. **The genetics and exploitation of heterosis in crops**. Madison: ASA / CSSA, 1999, p. 247-255.

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

SOUZA JUNIOR, C. L.; GARCIA, A. A. F. Índice para a seleção precoce em espécies de propagação assexuada. In: CONGRESSO NACIONAL DE GENÉTICA, 42., 1996, Caxambú: **Anais...** Ribeirão Preto: Sociedade Brasileira de Genética, 1996. Resumos.

SOUZA, A. C.; SILVA, S. E. L. Melhoramento de fruteiras da Amazônia. In: SIMPÓSIO BRASILEIRO DE FRUTEIRAS, 2. Viçosa: UFV, DFT, 2000. p. 49-55.

SOUZA, A. G. C.; SILVA, S. E. L.; SOUZA, N. R. Avaliação de progêneres de cupuaçuzeiro (*Theobroma grandiflorum*) em Manaus-AM. **Revista Brasileira de Fruticultura**, v. 20, n. 3, p. 307-312, 1998.

SOUZA, A. G.; BARCELOS, E.; CUNHA, R. N. V.; SOUSA, N. R. Parâmetros genéticos estimados em progêneres de cupuaçuzeiro. **Genetics and Molecular Biology**, v. 23, n. 3, p. 624, 2000.

SQUILLACE, A. E. Average genetic correlations among offspring from open pollinated forest trees. **Silvae Genetica**, v. 23, n. 5, p. 149-156, 1974.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. 2th. ed. New York: Mc Graw-Hill, 1980. 633 p.

STEIN, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 3., 1955, Berkeley. **Proceedings...** Berkeley: University of Berkeley, 1955. v. 1, p. 197-206.

- STRAM, D. O.; LEE, J. W. Variance components testing in longitudinal mixed effects setting. *Biometrics*, v. 50, p. 1171-1177, 1994.
- STRANDEN, I. J. *Robust mixed effects linear models with t distributions and application to dairy cattle breeding*. Madison: University of Wisconsin, 1996. Ph.D. Thesis.
- STRANDÉN, I.; LIDAUER, M. Solving large mixed linear models using preconditioned conjugate gradient iteration. *Journal of Dairy Science*, v. 82, p. 2779-2787, 1999.
- STREINER, D. L. Do you see what I mean? Indices of central tendency. *Canadian Journal of Psychiatry*, v. 45, n. 9, p. 833-836, 2000.
- STRINGER, J. K. Best linear unbiased prediction (BLUP) is a faster and more cost effective method of estimating breeding value in sugarcane. In: HYBRID BREEDING AND GENETICS OF FOREST TREES, 2000, Noosa. *Proceedings of QFRI/CRC-SPF Symposium*. Brisbane: QFRI / CRC / SPF, 2000. p. 505-509.
- STRINGER, J. K.; MCRAE, T. A.; COX, M. C. Best linear unbiased prediction as a method of estimating breeding value in sugar cane. In: WILSON, H.; HOGARTH, D.M.; CAMPBELL, S.; GARSIDE, G. (Ed.). *Sugarcane: research towards efficient and sustainable production*. Brisbane: CSIRO, 1996. p. 39-41.
- SUN, L.; HSU, J.S.J.; GUTTMAN, I.; LEONARD, T. Bayesian methods for variance component models. *Journal of the American Statistical Association*, v. 91, n. 434, p. 743-752, 1996.
- SWAIN, T. L.; GARDNER, R. A. W.; CHIAPERO, C. C. *Final report on ICFR Eucalyptus smithii trials in the summer rainfall region of South Africa*. Pietermaritzburg: ICFR, 2000. 13 p. (ICFR Bulletin Series, 14).
- SWAN, A. Across-breed genetic evaluation. In: HAMMOND, K.; GRASER, H-U.; Mc DONALD, C. A. *Animal breeding: the modern approach*. Sydney: Post Graduate Foundation in Veterinary Science, 1992. p. 111-120.
- SWAN, A.; KINGHORN, B. Design of crossbreeding programs. In: HAMMOND, K.; GRASER, H-U.; Mc DONALD, C. A. *Animal breeding: the modern approach*. Sydney: Post Graduate Foundation in Veterinary Science, 1992. p. 227-336.
- TALLIS, G. M. A selection index for optimum genotype. *Biometrics*, v. 18, p. 120-122, 1962.
- TALLIS, G. M. The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, v. 18, p. 342-353, 1962b.
- TALLIS, G. M. The use of a generalized multinomial distribution in the estimation of correlation in discrete data. *Journal of the Royal Statistical Society: Series B*, v. 24, p. 530-534, 1962a.

TASSELL, C. P. van; CASELLA, G.; POLLAK, E. J. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. *Journal of Dairy Science*, v. 78, p. 678-692, 1995.

TASSELL, C. P. van; VAN VLECK, L. D. van. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based covariance component inference. *Journal of Animal Science*, v. 74, p. 2586-2597, 1996.

TASSELL, C.P. van; VAN VLECK, L.D. van. **A manual for use of MTGSAM**: a set of FORTRAN programs to apply Gibbs sampling to animal models for variance component estimation. Whashington: USDA / ARS, 1995. 82p.

TEMPELMAN, R. J. Generalized linear mixed models in dairy cattle breeding. *Journal of Dairy Science*, v. 81, p. 1428-1444, 1998.

TEMPLETON, A. R.; READ, B. Inbreeding: one word, several meanings, much confusion. In: LOESCHKE, V. TOMIUK, J.; JAIN, S. K. (Ed). **Conservation Genetics**. Birkhäuser Verlag. Basel. p.91-105, 1994.

THOMPSON Jr., W. A. The problem of negative estimates of variance components. *Ann. Math. Stat.*, v. 33, p. 273-284, 1962.

THOMPSON, R. Design of experiments to estimate genetic parameters within populations. In: HILL, W.G.; MACKAY, T. (Ed.). **Evolution and animal breeding**. Edinburgh: CAB international, 1989. p. 169-74.

THOMPSON, R. Generalized linear models and applications to animal breeding. In: GIANOLA, D.; HAMMOND, K. (Ed.). **Advances in statistical methods for genetic improvement of livestock**. Berlin: Springer-Verlag, 1990. p. 312-328.

THOMPSON, R. Iterative estimation of variance components for non-orthogonal data. *Biometrics*, v. 25, p. 767-773, 1969.

THOMPSON, R. Recent developments in the estimation of variance components and their application to the estimation of genetic parameters. In: BARTON, R.A.; SMITH, W.C. (Ed.). WORLD CONGRESS ON SHEEP AND BEEF CATTLE BREEDING, Palmerston North, 1982. *Proceedings...* Palmerston North : Dunmore Press, 1982. v.1, p.217-224..

THOMPSON, R. Sire evaluation. *Biometrics*, v. 35, p. 339-353, 1979.

THOMPSON, R. The estimation of variance and covariance components when records are subject to culling. *Biometrics*, v. 29, p. 527-550, 1973.

THOMPSON, R.; MEYER, K. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science*, v. 15, p. 299-313, 1986.

TIETJEN, G. L.; MOORE, R. H.. On testing significance of components of variance in the unbalanced nested analysis of variance. *Biometrics*, n. 24, p. 423-429, 1968.

- TORGGLER, M. G. F. *Variação genética em progêneres dentro de procedências de Eucalyptus saligna Smith*. Piracicaba: ESALQ, 1987. 200 p. Tese Mestrado.
- TORO, M. A.; NIETO, B. A simple method for increasing the response to artificial selection. *Genetical Research*, v. 44, p.347-349, 1984.
- TORO, M. A.; PEREZ-ENCISO, M. Optimizing selection under restricted inbreeding. *Genet. Sel. Ev.*, v. 22, p. 93-107, 1990.
- TORO, M. A.; PRUNONOSA, D. V. The use of prior information in the estimation of heritability by parent-offspring regression. *Genetique, Selection, Evolution*, v. 16, p. 177-184, 1984.
- TORO, M. A.; SILIO, L. An alternative to restricted BLUP based on mate selection. *Livestock Production Science*, v. 32, p. 181-187, 1992.
- TORO, M. A.; SILIO, L.; PEREZ-ENCISO, M. A note on the use of mate selection in closed MOET breeding schemes. *Animal Production*, v. 53, p. 403-406, 1991.
- TUKEY, J. W. Bias and confidence in not-quite large samples. *Ann. Math. Statist.*, v. 29, p. 614, 1958.
- TUKEY, J. W. One degree of freedom for non-additivity. *Biometrics*, v.5, p.232-242, 1949.
- TURNER, H.N.; YOUNG, S.S.Y. *Quantitative genetics in sheep breeding*. Mc Millan of Australia, Melbourne, 1969, 332p.
- UGARTE, E.; ALENDA, R.; CARABAÑO, M. J. Fixed or random contemporary groups in genetic evaluations. *Journal of Dairy Science*, v. 75, n. 1, p. 269-278, 1992. v.61 , p. 1-26, 1999.
- VALEN, L. van. The statistics of variation. *Evolution Theory*, v. 4, p. 33-43, 1978.
- VALENTE, J. M. G. P. *Geomatemática*: lições de geoestatística. 2. ed. Ouro Preto: Fundação Gorceix, 1989. v. 3.
- VALOIS, A. C. C.; CORREA, M. P. F.; VASCONCELLOS, M. E. C. Estudos de caracteres correlacionados com a produção de amêndoa seca no guaranazeiro. *Pesquisa Agropecuária Brasileira*, v. 14, n. 2, p. 175-179, 1979.
- VALOIS, A. C. C.; NASCIMENTO, J. C. Estimativas de parâmetros genéticos em cacaueiros sem a utilização de testes de progêneres. *Pesquisa Agropecuária Brasileira*, v. 21, n. 9, p. 965-970, 1986.
- VARONA, L. *Aplicaciones del muestreo de Gibbs en modelos de genética cuantitativa*: análisis de un caso de heterogeneidad de varianzas. Zaragoza: Universidad de Zaragoza, 1994. PhD. Thesis.

VARONA, L.; MORENO, C.; GARCIA-CORTES, L. A.; ALTARRIBA, J. Estimación multicarácter de componentes de varianza y covarianza en vacuno lechero mediante muestreo de Gibbs. *Revista Portuguesa de Zootecnia*, v. 1, p. 185-195, 1994.

VASCONCELLOS, M. E. C.; GONÇALVES, P. S.; PAIVA, J. R.; VALOIS, A. C. C. Métodos de estimação do coeficiente de repetibilidade no melhoramento da seringueira. *Pesquisa Agropecuária Brasileira*, v. 20, n. 4, p. 433-437, 1985.

VENCOVSKY, R. Amostragem genética em populações naturais. *Silvicultura*, São Paulo, v. 41, p. 95-96, 1986.

VENCOVSKY, R. Análise da variância de freqüências alélicas. *Revista Brasileira de Genética*, v.15, n.3, supl., p. 53-60, 1992.

VENCOVSKY, R. Effective size of monoecious populations submitted to artificial selection. *Brazilian Journal of Genetics*, v. 1, n. 3, p. 181-191, 1978.

VENCOVSKY, R. Genética quantitativa. In: KERR, W. E. (Ed.). *Melhoramento e genética*. São Paulo: Melhoramentos, 1969. p. 17-38.

VENCOVSKY, R. Herança quantitativa. In: PATERNIANI, E.; VIEGAS, G. P. (Ed.) *Melhoramento e produção de milho*. 2.ed. Campinas: Fundação Cargill, 1987a. v. 1, p. 137-214.

VENCOVSKY, R. Preservação e genética de populações. In: ENCONTRO SOBRE RECURSOS GENÉTICOS, 1988. Jaboticabal. *Anais...* Jaboticabal: FCAVJ / UNESP. 1988, p. 67-74.

VENCOVSKY, R. Repetibilidade. In: VENCOVSKY, R. *Princípios de genética quantitativa*. Piracicaba: ESALQ/USP, 1972. p. 47-52.

VENCOVSKY, R. Tamanho efetivo populacional na coleta e preservação de germoplasma de espécies alógamas. *IPEF*, Piracicaba, n. 35, p. 7 9-84. 1987b.

VENCOVSKY, R. *Tamanho efetivo em populações submetidas à seleção: sexos separados*. Piracicaba: ESALQ/USP, 1978. p. 282-287. Relatório Científico do Departamento de Genética.

VENCOVSKY, R.; BARRIGA, P. *Genética biométrica no fitomelhoramento*. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486 p.

VENCOVSKY, R.; CROSSA, J. Variance effective population size under mixed self and random mating with applications to genetic conservation of species. *Crop Science*, v. 39, p. 1282-1294, 1999.

VENCOVSKY, R.; DIAS, C. T. S.; DEMETRIO, C. G. B.; LEANDRO, R. A.; PIEDADE, S. M. S. Reamostragem por bootstrap na estimação de parâmetros baseados em marcadores genéticos. In: BANDELL, G.; VELLO, N. A.; MIRANDA FILHO, J. B. (Ed.). *Genética biométrica vegetal: Anais*. Piracicaba: ESALQ/USP, 1997, p. 59-72.

- VENCOVSKY, R.; OSMANY, J. D.; RICARDO, Y. *Um modelo genético aplicado à análise de dados de leite em gado bovino*. Piracicaba: ESALQ/USP, 1970. p. 130-136. Relatório Científico do Departamento de Genética.
- VERBYLA, A. P.; CULLIS, B. R.; KENWARD, M. G.; WELHAM, S. J. The analyses of designed experiments and longitudinal data using smoothing splines. *Journal of the Royal Statistical Society: Series C*, v. 48, p. 269-311, 1999.
- VERRIER, E.; COLLEAU, J. J.; FOULLEY, J. L. Effect of mass selection on the within-family genetic variance in finite populations. *Theoretical and Applied Genetics*, v. 77, p. 142-148, 1989.
- VERRIER, E.; COLLEAU, J. J.; FOULLEY, J. L. Long-term effects of selection based on the animal model BLUP in a finite population. *Theoretical and Applied Genetics*, v. 87, p. 446-454, 1993.
- VERRYN, S.; FIELD, C.; GARCIA, J.; GRIFFIN, R. A discussion on the relationship between heritabilities and genetic correlations and the standard errors of these parameters with a case study example of GEI in *E. grandis* over two sites in South Africa and one site in Uruguay. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCALYPT, 1997, Salvador. *Proceedings...* Colombo: EMBRAPA. Centro Nacional de Pesquisa de Florestas, 1997. v. 1, p. 43-49.
- VIEIRA, S. *Estatística experimental*. São Paulo: Atlas, 1999, 185 p.
- VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.; ALVAREZ, V. H.; SCHAEFER, C. E. *Tópicos em ciência do solo*. Viçosa: Sociedade Brasileira de Ciência do Solo, 2000, p. 1-54.
- VIGNERON, P. Crédation et amélioration de variétés hybrides d'eucalyptus au Congo. In: IUFRO CONGRESS ON INTENSIVE FORESTRY: THE ROLE OF EUCALYPTUS, Durban, 1992. *Proceedings*. Durban: SAFRI, 1991, p.345-353.
- VIGNERON, P. Crédation et amélioration de variétés hybrides d'eucalyptus au Congo. *Bois et Forêts des Tropiques*, v. 234, n. 4, p. 29-42, 1992.
- VILLALOBOS, N. L.; GARRICK, D. J. Genetic parameters estimates for lamb survival in Romney sheep. *New Zealand Society of Animal Production*, v. 29, p. 41, 1999.
- VISSCHER, P. M.; GODDARD, M. E. Fixed and random contemporary groups. *Journal of Dairy Science*, v. 76, n. 5, p. 1444-1454, 1993.
- VISSCHER, P. M.; HILL, W. G. Heterogeneity of variance and dairy cattle breeding. *Animal Production*, v. 55, p. 321-329, 1992.
- VISSCHER, P. M.; THOMPSON, R.; HILL, W. G. Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livestock Production Science*, v. 28, p. 273-290, 1991.

VLECK, L. D. van. Estimation of heritability of threshold characters. *Journal of Dairy Science*, v. 55, n. 2, p. 218-225, 1972.

VLECK, L. D. van. **Selection index and introduction to mixed model methods**. Boca Raton: CRC Press, 1993a. 512 p.

VLECK, L.D. van. Estimation of nonadditive genetic variances for a total-merit model including maternal effects. *Journal of Animal Science*, v. 71, p. 2006-2010, 1993b.

VLECK, L.D. van; POLLAK, E. J.; OLTENACU, E.A. B. **Genetics for the animal sciences**. New York : W.H. Freeman, 1987. 391 p.

VOLKER, P. W. Evaluation of *Eucalyptus nitens* x *E. globulus* for commercial forestry. In: POTTS, B.; VOLKER, P.; BORRALHO, N.M.G.(Ed.). **Eucalypt plantations**: improving fibre yield and quality, Hobart, 1995. **Proceedings**. Hobart: CRC-IUFRO, 1995, p. 222-225.

VOLKER, P. W.; DEAN, C. A.; TIBBITS, W. N.; RAVENWOOD, I. C. Genetic parameters and gains expected from selection in *Eucalyptus globulus* in Tasmania. *Silvae Genetica*, Frankfurt, v. 39, n. 1, p. 18-28, 1990.

VONESH, E. F.; SHORK, M. A. Sample size analysis of repeated measurements. *Biometrics*, v. 42, n. 3, p. 601-610, 1986.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.*, v. 54, p. 426-482, 1943.

WALSH, J. L.; AHLBERS, J. H.; NILSON, E. N. Best approximation properties of the spline fit. *Journal of Mathematical Mechanics*, v.11, p. 225-234, 1962.

WALYARO, D. J.; VOSSEN, H. A. M. van der. Early determination of yield potential in arabica coffee by applying index selection. *Euphytica*, v. 8, p. 465-472, 1979.

WANG, C. S.; GIANOLA, D.; SORENSEN, D. A.; JENSEN, J.; CHRISTENSEN, A.; RUTLEDGE, J. J. Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis. *Theor. Appl. Genet.* v. 88, p. 220, 1994a.

WANG, C. S.; RUTLEDGE, J. J.; GIANOLA, D. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genetics Selection Evolution*, v. 26, p. 91, 1994b.

WANG, C. S.; RUTLEDGE, J. J.; GIANOLA, D. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genetics Selection Evolution*, v. 25, p. 41, 1993.

WEEKS, D. L.; WILLIAMS, D. R. A note on the determination of connectedeness in an N-way cross classification. *Technometrics*, v. 6, p. 319-324, 1964.

- WEI, M.; VAN DER WERF, J. H. J. Animal model estimation of additive and dominance variances in egg production traits of poultry. *Journal of Animal Science*, v. 71, p. 57, 1993.
- WEI, R. P. Loss of genetic diversity following selection from populations with a family structure. *Silvae Genetics*, v. 45, n. 2, p. 153-159, 1996.
- WEI, R. P.; LINDGREN, D. Effective family number following selection with restrictions. *Biometrics*, v. 52, p. 198-208, 1996.
- WEI, R. P.; LINDGREN, D. Optimal family contributions and a linear approximation. *Theoretical Population Biology*, v. 48, p. 318-332, 1995.
- WEI, R. P.; LINDGREN, D. Selection effects on diversity and genetic gain. *Silva Fennica*, v. 25, p. 229-234, 1991.
- WEI, X ; BORRALHO, N. M. G. A simple model to describe age trends in heritability in short rotation tree species. In: DIETERS, M. J.; MATHESON, A. C.; NIKLES, D. G.; HARWOOD, C.; WALKER, S.M. (Ed.). *Tree improvement for sustainable tropical forestry*. Caloundra: QFRI / IUFRO, 1996. p. 178-181.
- WEI, X.; BORRALHO, N. M. G. Genetic parameters for growth traits in four-year-old *Eucalyptus urophylla* progeny trial. In: POTTS, B.; VOLKER, P.; BORRALHO, N.M.G.(Ed.). *Eucalypt plantations*: improving fibre yield and quality, Hobart, 1995. *Proceedings*. Hobart: CRC-IUFRO, 1995, p. 252-254.
- WEIGEL, K. A.; GIANOLA, D. A computationally simple bayesian method for estimation of heterogeneous within-herd phenotypic variances. *Journal of Dairy Science*, v. 76, p. 1455-1465, 1993.
- WEINBERG, W. Further contributions to the theory of inheritance. *Arch. Rassen. Gesellschaftsbiol.*, v. 7, p. 35-49, p. 440-460, 1910.
- WEINBERG, W. On the laws of inheritance in man. I. General Part. Z. Indukt. Abstammungs Vererbungs l., v. 1, p. 377-392, 440-460, 1909.
- WEIR, B. S. *Genetic data analysis*. Sunderland: Sinauer Associates, Inc., 1990. 377 p.
- WEIR, B. S.; COCKERHAM, C. C. Estimating F statistics for the analysis of population structure. *Evolution*, v. 38, n. 8, p. 1358-1370, 1984.
- WEIR, B. S.; COCKERHAM, C. C. Two-locus theory in quantitative genetics. In: POLLAK, E.; KEMPTHORNE, O.; BAILEY, I. (Ed.). *Proceedings International Conference on Quantitative Genetics*. Ames: Iowa State University Press, 1977, p. 247-269.
- WELHAM, S. J.; THOMPSON, R. Likelihood ratio tests for fixed models terms using residual maximum likelihood. *Journal of the Royal Statistical Society, Series B*, v. 59, p. 701-719, 1997.

WERF, J. H. J. van der; BOER, I. J. M. Estimation of additive genetic variance when base populations are selected. *Journal of Animal Science*, v. 68, p. 3124-3132, 1990.

WESTELL, R. A.; QUAAS, R. L.; VLECK, L.D. van. Genetic group in an animal model. *Journal of Dairy Science*, v. 71, p. 1310-1318, 1988.

WHITE, A. G; ALPASH, P. A.; WESKETT, R. H.; BREWER, L. R. Heritability of fruit shape in pears. *Euphytica*, v. 112, n. 1, p. 1-7, 2000.

WHITE, I. M. S.; THOMPSON, R.; BROTHERSTONE, S. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy science*, v. 82, p. 632-638, 1999.

WHITE, T. Genetic parameters estimation and breeding value predictions: issues and implication in tree improvement programs. In: DIETERS, M. J.; MATHESON, A. C.; NIKLES, D. G.; HARWOOD, C. E.; WALKER, S. M. *Tree improvement for sustainable tropical forestry*. Proceedings. Caloundra: QFRI / IUFRO, 1996. p. 110-117.

WHITE, T. L.; HODGE, G. *Predicting breeding values with application in forest tree improvement*. London: Kluwer, 1989. 367 p.

WILCOX, M. D. Genetic improvement of eucalyptus in New Zealand. *New Zealand Journal of Forest Science*, v. 10, p. 343-359, 1980.

WILCOX, M. D. Inbreeding depression and genetic variances estimated from self and cross – pollinated families of *Pinus radiata*. *Silvae Genetica*, v. 32, n. 3, p. 89-96, 1983.

WILLHAM, R. L. The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics*, v. 19, p. 18-27, 1963.

WILLHAM, R. L. The role of maternal effects in animal breeding. III. Biometrical aspects of maternal effects in animals. *Journal of Animal Science*, v. 35, n. 6, p. 1288-1293, 1972.

WILLIAMS, C. G.; SAVOLAINEN, O. Inbreeding depression in conifers. *Forest Science*, v. 41, n. 2, p. 1-20. 1996.

WILLIAMS, E.; MATHESON, A. C. *Experimental designs and analysis for use in tree improvement*. Victoria: CSIRO, 1994. 174 p.

WILLIAMS, J. S. The evaluation of a selection index. *Biometrics*, v. 18, p. 375-393, 1962.

WILLIS, M. B. *Dalton's introduction to practical animal breeding*. 3th ed. Oxford: Blackwell Science, 1991. 159 p.

- WILLYAMS, D.; MOOLJHUZEN, P.; VOLKER, P.; RAYMOND, C.; CHANDLER, S. F. Micropropagation of juvenile *Eucalyptus regnans*. In: MASS PRODUCTION TECHNOLOGY FOR GENETICALLY IMPROVED FAST GROWING FOREST TREE SPECIES. 1992b. Bordeaux. *Proceedings...* Nangis: Association Forêt Cellulose, 1992b. p. 415-419.
- WILLYAMS, D.; WHITEMAN, P.; CAMERON, J.; CHANDLER, S. F. Inter-and intra-family variability of rooting capacity in micropropagated *Eucalyptus globulus* and *Eucalyptus nitens*. In: MASS PRODUCTION TECHNOLOGY FOR GENETICALLY IMPROVED FAST GROWING FOREST TREE SPECIES, 1992. Bordeaux. *Proceedings....* Nangis: Association Forêt Cellulose, 1992a. p. 177-181.
- WILSON, P. J. The development of new clones of *Eucalyptus globulus* and *E. globulus* hybrids by stem cuttings propagation. In: MASS PRODUCTION TECHNOLOGY FOR GENETICALLY IMPROVED FAST GROWING FOREST TREE SPECIES, Bordeaux. *Proceedings...* Nangis: Association Forêt Cellulose, 1992. p. 379-386.
- WOLFINGER, R. D.; FEDERER, W. T.; CORDERO-BRANA, O. Recovering information in augmented designs using SAS PROC GLM and PROC MIXED. *Agronomy Journal*, v. 89, p. 856-859, 1997.
- WOOLASTON, R. R.; JARVIS, S. F. The importance of breeding objectives in forest tree improvement. In: : POTTS, B.; VOLKER, P.; BORRALHO, N.M.G.(Ed.). *Eucalypt plantations*: improving fibre yield and quality, Hobart, 1995. *Proceedings*. Hobart: CRC-IUFRO, 1995, p. 184-188.
- WOOLASTON, R. R.; KANOWSKI, P. J.; NIKLES, D. G. Genetic parameters estimates for *Pinus caribaea* var. *hondurensis* in Coastal Queensland, Australia. *Silvae Genetica*, v. 39, n. 1, p. 21-28, 1990.
- WOOLIAMS, J. A.; WRAY, N. R.; THOMPSON, R. Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genetical Research*, v. 62, p. 231-242, 1993.
- WOOLONS, R. C. Importance of experimental design and plot maintenance in forest field experiments in Australasia. *Australian Forest Research*, v. 10, p. 71-82, 1980.
- WRAY, N. R.; WOOLIAMS, J. A.; THOMPSON, R. Prediction of rates of inbreeding in populations undergoing index selection. *Theoretical and Applied Genetics*, v. 87, p. 878-892, 1994.
- WRICKE, G.; WEBER, W. E. *Quantitative genetics and selection in plant breeding*. Berlin: Walter de Gruyter. 1986. 406 p.
- WRIGHT, J. M.; FREELAND, F. D. *Plot size and experimental efficiency in forest genetics research*. East Lansing: Michigan State University, 1960. 27 p. (Agr. Exp. Stat. Tech. Bull, 280).

- WRIGHT, J. W. *Introduction to forest genetics*. New York: Academic Press, 1976. 463 p.
- WRIGHT, J. W.; COCKERHAM, C. C. Selection with partial selfing. I. Mass selection. *Genetics*, v. 109, p. 585-597, 1985.
- WRIGHT, J. W.; COCKERHAM, C. C. Selection with partial selfing. II. Family selection. *Crop Science*, v. 26, p. 261-268, 1986.
- WRIGHT, S. Systems of mating. *Genetics*, v. 6, p. 111-178, 1921.
- WRIGHT, S. The genetical structure of populations. *Annals of Eugenics*, London, v. 15, p. 323-354, 1951.
- WYK, G. van. Early growth results in a diallel progeny test of Eucalyptus grandis. 1. A field study. *Silvae Genetica*, v. 25, p. 126-132, 1976.
- WYK, G. van. Genetic improvement of timber yield and wood quality in Eucalyptus grandis. I. Genetic parameters of growth characteristics. *South African Forestry Journal*, v. 153, p. 1-11, 1990.
- WYK, G. van. Inbreeding effects in Eucalyptus grandis in relation to the degree of relatedness. *South African Forestry Journal*, v. 116, p. 60-63, 1981.
- YATES, F. A new method of arranging variety trials involving a large number of varieties. *Journal of Agricultural Sciences*, v. 26, p. 424-455, 1936.
- YATES, F. The recovery of inter-block information in balanced incomplete block designs. *Annals of Eugenics*, v. 10, p. 317-325, 1940.
- YAZDI, M. H.; JOHANSSON, K.; GATES, P.; NASHOLM, A. Bayesian analysis of birth weight and litter size in Baluchi sheep using Gibbs sampling. *Journal of Animal Science*, v. 77, p. 533-540, 1999.
- YOUNG, S. S. Y. A further examination of relative efficiency of three methods of selection for genetic gains under less restricted conditions. *Genetical Research*, v. 2, p. 106-121, 1961.
- YOUNG, S. S. Y.; SKAVARIL, R. V. Computer simulations of within family selection in finite populations. *Theoretical and Applied Genetics*, v. 48, p. 45-51, 1976.
- YOUNG, S. S. Y.; TALLIS, G. M. Performance index for lifetime production. *Journal of Animal Science*, p. 506-509, 1961.
- YOUNG, S. S. Y.; WEILER, H. Selection for two correlated traits by independent culling levels. *Journal of Genetics*, v. 57, p. 329-338, 1961.
- YUAN, J. Y. *Iterative methods for the generalized least-squares problem*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 1994. 139 p. (Informes de Matemática, Série F-070/94).
- ZAHL, S. Jack-Knifing an index of diversity. *Ecology*, v. 58, p. 907-913, 1977.

ZELLNER, A. *An introduction to Bayesian inference in econometrics*. New York: J. Wiley and Sons, 1971. 430 p.

ZIMMERMAN, D. I.; HARVILLE, D. A. A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics*, v. 47, p. 223-239, 1991.

**Índice Remissivo de
Espécies Perenes e Anuais**

Índice Remissivo de Espécies Perenes e Anuais

- Abacaxi **835**
Acácia *auriculiformis* **732**
Acácia *mangium* **732**
Acácia negra **478, 618, 619, 620, 649, 732, 734, 750**
Açaí **750**
Acerola **134, 141, 146, 202, 203, 204, 205, 478, 651, 654, 739**
Alfafa **345**
Algodão **790**
Alho **835**
Banana **835**
Batata **345, 835**
Batata-doce **835**
Berinjela **790**
Beterraba **345**
Braquiária **345, 752**
Cacau **128, 168, 472, 478, 651, 653, 654, 728, 735, 739, 751**
Café arábica **345, 478, 651, 653, 728, 735, 751, 754, 761, 762, 767, 768, 769, 775**
Café robusta **76, 158, 651, 728, 736, 739, 751, 762**
Caju **335, 336, 478, 653, 738, 739, 848, 850**
Cana-de-açúcar **829, 835**
Capim-elefante **345, 835**
Centeio **345**
Chá-da-Índia **734, 735, 835**
Coco **160, 168, 207, 478, 654, 655, 728, 736, 737, 739, 751**
Cupuaçú **168, 206, 478, 652, 654, 739, 750**
Dendê **478, 650, 653, 728, 737, 739, 751**
Erva-mate **168, 173, 175, 199, 376, 377, 378, 478, 634, 636, 650, 652, 654, 655, 672, 678, 750, 846, 847, 886**
Eucalipto **241, 251, 268, 285, 290, 297, 309, 327, 358, 373, 374, 376, 456, 478, 483, 488, 517, 518, 520, 574, 645, 646, 647, 648, 654, 655, 708, 719, 728, 729, 730, 731, 739, 745, 746, 747, 748, 749, 750, 751, 790, 829**
Girassol **790**
Guaraná **168, 207, 478, 641, 653, 654**
Laranja **835**
Leucena **345, 734, 754, 761, 762**
Maçã **478, 650, 739**
Mandioca **829, 835**
Melancia **790**
Melão **790**
Milho **771, 788**
Morango **835**
Panicum **345, 751**
Pêra **651**

Pêssego **754, 761, 762**

Pinus **207, 290, 478, 519, 526, 530, 649, 652, 655, 678, 731, 732, 750**

Pupunha **478, 652, 737, 738, 750**

Quiabo **790**

Seringueira **168, 206, 207, 478, 570, 574, 575, 650, 652, 654, 655, 678, 728, 731, 732, 733, 734, 739, 750, 751, 787, 790, 852**

Soja **771**

Sorgo **790**