



**ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA
EM MISTURAS DE DENSIDADES NORMAIS: UMA
APLICAÇÃO EM GENÉTICA**

WALDEMAR PAMPLONA DA SILVA

2003

WALDEMAR PAMPLONA DA SILVA

**ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA
EM MISTURAS DE DENSIDADES NORMAIS: UMA
APLICAÇÃO EM GENÉTICA**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

Orientador

Dr. Eduardo Bearzoti

LAVRAS
MINAS GERAIS - BRASIL
2003

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Silva, Waldemar Pamplona da

**Estimadores de máxima verossimilhança em mistura de densidades normais:
uma aplicação em genética / Waldemar Pamplona da Silva. – Lavras : UFLA,
2003.**

60 p. : il.

**Orientador: Eduardo Bearzoti.
Dissertação (Mestrado) – UFLA.
Bibliografia.**

**1. Estimador de máxima verossimilhança. 2. Mistura de densidade normal. 3.
Herança genética. 4. Razão de Verossimilhança. I. Universidade Federal de
Lavras. II. Título.**

CDD-519.5

WALDEMAR PAMPLONA DA SILVA

**ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA
EM MISTURAS DE DENSIDADES NORMAIS: UMA
APLICAÇÃO EM GENÉTICA**

**Dissertação apresentada à Universidade
Federal de Lavras como parte das
exigências do Programa de Pós-graduação
em Agronomia, área de concentração em
Estatística e Experimentação Agropecuária,
para obtenção do título de "Mestre".**

APROVADA em 31 de março de 2003

Daniel Furtado Ferreira, Dr.

UFLA

Wilson Roberto Maluf, PhD.

UFLA

Júlio Sílvio de Sousa Bueno Filho, Dr.

UFLA



Prof. Eduardo Bearzoti, Dr.

DEX/UFLA

(Orientador)

LAVRAS

MINAS GERAIS - BRASIL

A Deus,

por sempre iluminar meu caminho.

Aos meus pais,

José Pamplona da Silva e Maria de Lourdes Augusta (†),

pelo exemplo de vida.

Dedico

À minha esposa, Rozana; às minhas filhas, Andrezza e Greicy, pela compreensão da ausência, pelo carinho e apoio; ao meu filho, Christian (†), que por poucos meses nos deixou lição de vida; às minha irmãs, Aparecida, Maria das Graças, Terezinha e Roseli, e aos demais familiares pelo convívio, incentivo e apoio em todos os momentos.

Ao professor Eduardo Bearzoti pela amizade, convívio e grande contribuição à minha vida profissional.

Ofereço

AGRADECIMENTOS

A Deus, por tudo.

Ao professor Eduardo Bearzoti – Dr., não só pela orientação e pelos valiosos ensinamentos, mas também pela pessoa educada, prestativa, sincera e amiga que realmente é. Pelo profissionalismo, pela responsabilidade, pelos resultados pertinentes encontrados, mostrando sua grande contribuição à ciência. Por me permitir adquirir uma visão holística da pesquisa científica. E, sobretudo, pelo convívio e amizade.

À Universidade Federal de Lavras (UFLA), ao Departamento de Ciências Exatas (DEX), em especial ao Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, pela oportunidade de realização do curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela bolsa de estudo concedida.

Ao Centro Federal de Educação Tecnológico de Uberaba-MG pela liberação concedida.

Aos professores Daniel Furtado Ferreira e Wilson Roberto Maluf pelo auxílio e sugestões.

Aos professores membros da banca examinadora, Eduardo, Daniel, Júlio e Wilson, pela colaboração e sugestões.

Aos professores do departamento de Ciências Exatas pelos conhecimentos transferidos.

A todos os funcionários do departamento de Ciências Exatas pela competência e prestabilidade.

Aos meus colegas de curso, Marcelo Cirillo, José Airton, Washington, Otaviano, Lúcia, Luis Alberto, Élcio, João Marcos, Ana, Alessandra e Ermelino, pelo companheirismo e carinho de verdadeiros irmãos.

Aos meus colegas de trabalho, professores e técnicos administrativos do CEFET-Uberaba MG, pelo apoio e incentivo a esta caminhada.

A todos aqueles que direta ou indiretamente contribuíram para a realização deste trabalho.

BIOGRAFIA

Waldemar Pamplona da Silva, filho de José Pamplona da Silva e Maria de Lourdes Augusta (†), nasceu na cidade de Conquista, Estado de Minas Gerais, em 09 de maio de 1955.

Concluiu o curso de graduação em Matemática nas Faculdades Integradas Santo Tomás de Aquino (FISTA-Uberaba), em dezembro de 1980.

Concluiu o curso de graduação em Ciências com Habilitação Plena em Física, na UNIFRAN-Franca – União das Faculdades Francanas, em dezembro de 1985.

Concluiu o curso de Especialização em Matemática “Latu-Sensu” na Faculdade de Filosofia, Ciências e Letras de Patrocínio, em 1993.

Foi professor de matemática da rede estadual de ensino em Conquista e Uberlândia MG, no período de 1977 a 1993.

Foi professor de matemática na Escola Agrotécnica Federal de Januária – MG, no período de abril de 1993 a novembro de 1995.

É professor na área de ciências exatas do SEFET-Uberaba –MG.

Em março de 2001, iniciou o curso de mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, na Universidade Federal de Lavras (UFLA), concluindo-o em março de 2003.

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 Estudos de Herança em Características Contínuas	3
2.2 Herança Poligênica	5
2.3 Herança Monogênica	11
2.3.1 Metodologia de Arias et al. (1994)	12
2.3.2 Metodologia Apresentada por Souza Sobrinho (1998).....	14
2.4 Modelos de Misturas	16
2.4.1 Métodos Numéricos Alternativos	19
3 METODOLOGIA.....	21
3.1 Construção do Modelo Genético.....	22
3.2 Estimação	26
3.2.1 Derivada Parcial Primeira em Relação a Um Componente Qualquer de Média	28
3.2.2 Derivada Parcial Primeira em Relação a Um Componente Qualquer da Variância	29
3.2.3 Derivada Parcial Segunda em Relação a Dois Componentes Quaisquer de Média	30
3.2.4 Derivada Parcial Segunda em Relação a Um Componente Qualquer de Média e um Componente Qualquer de Variância	32
3.2.5 Derivada Parcial Segunda Relativa a Dois Componentes Quaisquer de Variância	34
3.2.6 Função Escore do Modelo Genético	38
3.3 Sub-Modelos	39

3.4 Testes de Hipóteses	43
3.5 Software Estatístico Para Estudos de Herança	45
4 EXEMPLO DE APLICAÇÃO.....	47
4.1 Testes de Hipóteses de Modelos Genéticos Hierarquizados de Interesse.....	47
5 CONCLUSÕES.....	51
REFERÊNCIAS BIBLIOGRÁFICAS	52
ANEXOS	55

RESUMO

SILVA, Waldemar Pamplona da. **Estimadores de máxima verossimilhança em misturas de densidades normais: uma aplicação em genética.** LAVRAS: UFLA, 2003. 60p. (Dissertação – Mestrado em Agronomia/Área de concentração em Estatística e Experimentação Agropecuária)*

Estudos de herança genética são relativamente comuns no melhoramento de plantas, e particularmente úteis em certas características contínuas, para elucidar se a herança é devida à ação de um único gene ou predominantemente à ação de muitos genes de efeitos menores. Para tal, fazem, de maneira geral, uso de diferentes populações, as populações genitoras contrastantes P_1 e P_2 , F_1 (cruzamento entre indivíduos de P_1 e P_2), RC_{11} (cruzamento entre indivíduos de F_1 e P_1), RC_{12} (cruzamento entre indivíduos de F_1 e P_2) e F_2 (autocruzamento de indivíduos de F_1). Todavia, os valores genotípicos dos homocigotos e do heterocigoto são geralmente estimados fazendo o uso apenas das gerações P_1 , P_2 e F_1 . Este trabalho teve como objetivos: a obtenção de estimadores fazendo o uso destas seis populações, pelo método da máxima verossimilhança; propor critérios para testar hipóteses de interesse e ilustrar a metodologia fazendo o uso de um conjunto de dados reais. Para tanto, admitiu-se que a cada classe genotípica era associada uma distribuição normal com média e variâncias específicas, de maneira que as distribuições relativas às populações RC_{11} , RC_{12} e F_2 correspondem a misturas de distribuições normais. Os modelos genéticos foram estabelecidos e os estimadores obtidos, resultando em expressões recursivas. Para ilustração, utilizou-se um conjunto de dados reais para estudo do fenômeno da partenocarpia em abobrinha. Concluiu-se que esta característica é de herança monogênica com ação de dominância.

* Comitê de Orientação: Dr. Eduardo Bearzoti - UFLA (Orientador), Dr. Daniel Ferreira Furtado - UFLA, PhD. Wilson Roberto Maluf – UFLA.

ABSTRACT

SILVA, Waldemar Pamplona da. **Maximum likelihood estimators in mixtures of normal densities: an application in genetics.** Lavras: UFLA, 60p. (Dissertation - Master in Agronomy/Major in Statistics and Agricultural Experimentation).

Studies of genetic inheritance are relatively common in plant breeding, and specially useful in continuous traits, to clarify the nature of the inheritance, whether a single gene is involved and/or polygenes are present. Such studies generally make use of populations such as inbred parents, P_1 and P_2 , F_1 (crossing among parents), BC_{11} (crossing of F_1 and P_1), BC_{12} (crossing of F_1 and P_2), and F_2 (selfing of F_1 individuals). However, the genotypic values of homozygotes and heterozygotes are generally estimated using only P_1 , P_2 and F_1 information. This work aimed at: obtaining estimators that make use of all populations, via maximum likelihood; proposing criteria to test relevant hypotheses; and illustrating the approach with an actual data set. The statistical model considered a normal density with a specific mean for each genotype of the putative major gene, and a particular variance according to the population. Therefore, densities of BC_{11} , BC_{12} and F_2 populations were mixtures of normal distributions. Iterative procedures were needed to yield the maximum likelihood estimates. The illustration was a study of the inheritance of partenocarpy in *Cucurbita pepo* L.. Results suggested that this trait is governed by a major gene showing dominance.

Guidance Committee: Dr. Eduardo Bearzoti - UFLA (Adviser), Dr. Daniel Ferreira Furtado – UFLA and PhD. Wilson Roberto Maluf – UFLA.

1 INTRODUÇÃO

No melhoramento genético em culturas de uma maneira geral, o conhecimento acerca do controle genético (magnitude de efeitos e tipo de ação gênica) de características agronômicas de interesse é um aspecto relevante, que permite determinar e otimizar métodos de melhoramento.

Quando a característica em questão é contínua, é de especial interesse verificar se ocorre um gene de efeito maior, ou se a herança é essencialmente poligênica. Se há um gene de efeito maior, apesar de a característica ser contínua, então a variação dentro de cada classe genotípica se deve à ação de efeitos ambientais, somada ou não à ação de genes de menor efeito (poligenes).

Em geral, os estudos de herança têm considerado a estimação de parâmetros associados aos poligenes (componentes de média e de variância) e ao gene de efeito maior de maneira estanque, ou seja, sem a utilização de um modelo único e geral, simultaneamente. Além disso, no tocante ao gene de efeito maior, os valores genotípicos são estimados, em geral, utilizando apenas as gerações correspondentes às linhagens genitoras contrastantes (designadas por “P₁” e “P₂”) e ao cruzamento entre elas (geração “F₁”), embora outras gerações também estejam em geral presentes. Dentre estas destacam-se as gerações F₂ (autofecundação de indivíduos F₁), RC₁₁ (cruzamento entre indivíduos F₁ e P₁) e RC₁₂ (cruzamento entre indivíduos F₁ e P₂).

Face ao exposto, o presente trabalho teve por objetivos:

a) obtenção de estimadores de máxima verossimilhança de valores genotípicos relativos a um eventual gene de efeito maior em características contínuas, bem como dos componentes de média e de variância relativos aos poligenes, utilizando dados de gerações P₁, P₂, F₁, F₂, RC₁₁ e RC₁₂;

b) propor critérios de teste de hipóteses de interesse, como a da existência de um gene de efeito maior, de poligenes, bem como de interação alélica de dominância;

c) ilustrar a metodologia proposta com um conjunto de dados reais.

2 REFERENCIAL TEÓRICO

2.1 Estudos de herança em características contínuas

Estudos de herança são relativamente mais comuns e mais simples quando a característica em questão é qualitativa. Em tais casos, em geral testes de aderência são feitos, considerando diferentes hipóteses sobre a segregação, que refletem, por sua vez, diferentes freqüências esperadas em cada categoria.

Se a característica de interesse é contínua, então é freqüente admitir modelos da genética quantitativa, que se baseiam na existência de genes de pequeno efeito (poligenes) e pronunciada influência ambiental. Dispondo-se de dados de gerações típicas como as genitoras contrastantes P_1 (genitor 1) e P_2 (genitor 2), F_1 (cruzamento entre indivíduos de P_1 e P_2), F_2 (cruzamento entre indivíduos de F_1), RC_{11} (cruzamento entre indivíduos de P_1 e F_1) e RC_{12} (cruzamento entre indivíduos de P_2 e F_1), pode-se estimar componentes de variância utilizando a análise de variância, e componentes de média utilizando a metodologia do teste de escala conjunta, conforme descrito por Mather & Jinks (1984).

No entanto, em alguns estudos, apesar de a característica ser contínua, é do interesse do pesquisador indagar acerca da existência de um gene de efeito maior, associado ou não a poligenes. No Brasil, alguns estudos têm sido realizados com este enfoque.

Arias et al. (1996) estudaram a herança da resistência à doença mancha olho-de-rã, na cultura da soja, provocada por *Cercospora sojina*. Embora os autores tenham utilizado uma escala discretizada de 0 a 5, é claro que este tipo de avaliação busca representar uma escala contínua de resistência. Utilizando

diferentes cruzamentos entre cultivares de soja, foi possível identificar alelos de resistência às raças Cs-4 e Cs-15 do patógeno.

Souza Sobrinho (1998) relatou um estudo de herança da reação de resistência à raça 2 de *Meloidogyne incógnita* em *Capsicum annuum*, utilizando as cultivares 'Carolina Cayenne' e 'Agrônômico-8' como genitores. O autor ainda dispunha das gerações F₁, F₂, RC₁₁ e RC₁₂. As características avaliadas, embora discretas (número de galhas e número de massas de ovos), permitiram uma aproximação normal, e o autor pôde identificar um gene de efeito maior responsável pela resistência.

Em alface, a resistência ao nematóide das galhas foi estudada por Gomes (1999), utilizando um enfoque semelhante. O autor dispunha de gerações P₁, P₂, F₁ e F₂, além de famílias F₃. Também aqui foi possível inferir que a herança da resistência é monogênica.

Resende (1999) estudou a herança do teor de acilaçúcares em tomate, os quais conferem resistência a várias pragas. Os genitores consistiam de duas espécies de tomate, tendo sido obtidas as gerações F₁ e F₂. Por se tratar de uma concentração, a característica de interesse foi tipicamente contínua. Os resultados deram suporte à hipótese de herança monogênica, com dominância parcial no sentido do menor teor de acilaçúcares.

A herança da resistência ao vírus da mancha anelar em melancia foi estudada por Azevedo (2001). Além de dois genitores contrastantes, foram avaliadas plantas das gerações F₁, F₂, RC₁₁ e RC₁₂. Da mesma maneira como no trabalho de Arias et al. (1996), a característica resistência foi discretizada através de um sistema de notas. No estudo de Azevedo (2001), a hipótese de herança monogênica foi rejeitada, tendo a inferência se baseado nos componentes de média e de variância de poligenes, que sugeriram uma ação gênica predominantemente aditiva.

Freitas et al. (2002) reportaram um estudo sobre a herança do zingibereno em folhas de tomateiro, o qual também está relacionado à resistência a pragas. Foram utilizadas as gerações P₁, P₂, F₁ e F₂. Houve evidências da ação predominante de um gene de efeito maior, sendo o alelo de alto conteúdo parcialmente recessivo ao alelo de baixo conteúdo.

Quando a herança é poligênica, as metodologias mais consagradas para estudá-la são, possivelmente, a análise de variância (para estimar componentes de variância) e o teste de escala conjunto (para inferir sobre componentes de médias). Sendo a herança monogênica e a característica contínua (pelo menos aproximadamente), destacam-se as metodologias apresentadas por Arias et al. (1994) e por Souza Sobrinho (1998). Estas metodologias estão descritas a seguir, considerando cada tipo de herança em separado.

2.2. Herança Poligênica

Uma das metodologias mais consagradas para inferir sobre componentes de médias de características de herança poligênica é o chamado teste de escala conjunto, o qual está descrito em Mather & Jinks (1984) e em Ramalho et al. (1993).

O teste de escala conjunto é baseado no método dos quadrados mínimos ponderados. Por este método, utiliza-se um modelo linear:

$$Y = X\theta + \varepsilon \quad (1.1),$$

sendo Y o vetor de observações, com matriz de variâncias e covariâncias diagonal, representada por V ; X é a matriz de incidência, θ é o vetor de parâmetros, e ε é o vetor de resíduos.

Uma vez que V é diagonal no presente caso (covariâncias nulas mas variâncias possivelmente diferentes devido à existência de dados de diferentes

gerações , as quais apresentam variabilidades genéticas distintas), V^{-1} também o é, e assim a sua decomposição de Cholesky resulta em:

$$V^{-1} = LL',$$

sendo L uma matriz diagonal contendo, como elementos, as raízes quadradas dos elementos de V^{-1} , nas posições correspondentes. Pré-multiplicando (1.1) por L , resulta

$$LY = LX\theta + L\varepsilon \quad (1.2),$$

o qual também é um modelo linear, porém homocedástico e de variância igual a 1, pois:

$$V(LY) = LV(Y)L' = L(LL')^{-1}L' = LL^{-1}L^{-1}L' = I.$$

Utilizando o sistema de equações normais para estimar θ no modelo (1.2), e admitindo que LX tenha posto coluna completo, então:

$$\begin{aligned} \hat{\theta} &= [(LX)'(LX)]^{-1}(LX)'LY = \\ &[X'LLX]^{-1}X'LLY = (X'V^{-1}X)^{-1}X'V^{-1}Y. \end{aligned} \quad (1.3)$$

Os resíduos de (1.2) são estimados por:

$$Le = LY - LX\hat{\theta},$$

sendo que a forma quadrática

$$(Le)'(Le) = e'LL'e = e'V^{-1}e \quad (1.4),$$

sob normalidade, tem distribuição qui-quadrado com número de graus de liberdade igual ao tamanho da amostra, menos o número de parâmetros estimados.

A metodologia do teste de escala conjunto utiliza um modelo linear que descreve médias de diferentes gerações considerando parâmetros com clara interpretação genética. Assim, pode-se definir os componentes de média: μ , sendo esta uma constante de referência; $[a]$, sendo a soma dos efeitos aditivos dos poligenes; e $[d]$, definido como a soma dos efeitos de dominância dos poligenes. A metodologia é aqui ilustrada considerando as gerações $P_1, P_2, F_1,$

F_2 , RC_{11} e RC_{12} , mas outras gerações podem ser utilizadas (Mather & Jinks, 1984). Com aqueles componentes de média, tem-se que os valores fenotípicos dos indivíduos das gerações acima podem ser expressos por um modelo linear contendo a esperança da geração em questão, mais um desvio aleatório:

$$P_1 : Y_{1i} = \mu - [a] + \varepsilon_{1i}, \quad \varepsilon_{1i} \sim N(0, V_{P_1}), \quad i = 1, 2, \dots, n_1,$$

$$P_2 : Y_{2i} = \mu + [a] + \varepsilon_{2i}, \quad \varepsilon_{2i} \sim N(0, V_{P_2}), \quad i = 1, 2, \dots, n_2,$$

$$F_1 : Y_{3i} = \mu + [d] + \varepsilon_{3i}, \quad \varepsilon_{3i} \sim N(0, V_{F_1}), \quad i = 1, 2, \dots, n_3,$$

$$RC_{11} : Y_{4i} = \mu - \frac{[a]}{2} + \frac{[d]}{2} + \varepsilon_{4i}, \quad \varepsilon_{4i} \sim N(0, V_{RC_{11}}), \quad i = 1, 2, \dots, n_4,$$

$$RC_{12} : Y_{5i} = \mu + \frac{[a]}{2} + \frac{[d]}{2} + \varepsilon_{5i}, \quad \varepsilon_{5i} \sim N(0, V_{RC_{12}}), \quad i = 1, 2, \dots, n_5,$$

$$F_2 : Y_{6i} = \mu + \frac{[d]}{2} + \varepsilon_{6i}, \quad \varepsilon_{6i} \sim N(0, V_{F_2}), \quad i = 1, 2, \dots, n_6.$$

Assim, está-se admitindo variâncias diferentes nas gerações. Tais variâncias apresentam um componente ambiental e um componente genético que são definidos adiante, no item 3 (Metodologia). Por hora, basta considerar que, se amostras aleatórias são tomadas de tais gerações, calculando-se as médias de cada geração, é possível representá-las tomando como ponto de partida os modelos acima, por meio da relação matricial:

$$\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \\ \bar{Y}_4 \\ \bar{Y}_5 \\ \bar{Y}_6 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1/2 & 1/2 \\ 1 & 1/2 & 1/2 \\ 1 & 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ [a] \\ [d] \end{bmatrix} + \begin{bmatrix} \bar{\varepsilon}_1 \\ \bar{\varepsilon}_2 \\ \bar{\varepsilon}_3 \\ \bar{\varepsilon}_4 \\ \bar{\varepsilon}_5 \\ \bar{\varepsilon}_6 \end{bmatrix} \quad (1.5).$$

$$Y = X \cdot \theta + \varepsilon$$

quadrados mínimos ignora a distribuição de probabilidade de algumas gerações, como F_2 e retrocruzamentos, o que corresponde a misturas de densidades.

O teste de escala conjunto utiliza estimativas de componentes de variâncias. Estas, por sua vez, são em geral obtidas através da análise de variância. É muito freqüente, no melhoramento de plantas, a utilização de um delineamento experimental com controle local para avaliar plantas de diferentes gerações, considerando cada uma destas como um tratamento de efeito fixo. Se o delineamento for em blocos casualizados, tem-se a análise de variância representada na Tabela 1, exemplificada com as gerações P_1 , P_2 , F_1 , RC_{11} , RC_{12} e F_2 .

Algumas das variâncias dentro de cada geração, apresentadas na Tabela 1, possuem componentes genéticos particulares que, por enquanto, não vêm ao caso. O fato relevante aqui é o de que, igualando-se os quadrados médios às suas esperanças, obtêm-se as estimativas necessárias (método dos momentos) para a implementação do teste de escala conjunto.

Tabela 1 - Esquema de análise de variância utilizada em estudos de herança, avaliando-se gerações P₁, P₂, F₁, RC₁₁, RC₁₂, e F₂ em um delineamento em blocos ao acaso com r repetições.*

Causas de Variação	Graus de Liberdade	Esperanças dos Q. Médios
Bloco	(r - 1)	—
Gerações	5	—
Erro	5(r - 1)	$\sigma_d^2 + \bar{k} \sigma_e^2$ **
Dentro de P ₁	r(K ₁ - 1)	V _{P1}
Dentro de P ₂	r(K ₂ - 1)	V _{P1}
Dentro de F ₁	r(K ₃ - 1)	V _{F1}
Dentro de RC ₁₁	r(K ₄ - 1)	V _{RC11}
Dentro de RC ₁₂	r(K ₅ - 1)	V _{RC12}
Dentro de F ₂	r(K ₆ - 1)	V _{F2}

* K_j = número de plantas por parcela na geração j, admitido constante em todas as repetições. ** A esperança aqui é aproximada. σ_d^2 é a variância média dentro das gerações, \bar{K} é a média harmônica dos K_j e σ_e^2 é componente da variância ambiental entre parcelas.

É importante ressaltar, contudo, que se um delineamento experimental foi utilizado, conforme exemplificado na Tabela 1, então passa a existir uma variância ambiental dentro de parcelas, e outra entre parcelas, quantificada pelo erro experimental. Assim, as variâncias das médias de gerações passam a ser:

$$P_1 : V(\bar{Y}_{1.}) = \frac{V_{P_1}}{rK_1} + \frac{\sigma_e^2}{r}$$

$$P_2 : V(\bar{Y}_{2.}) = \frac{V_{P_2}}{rK_2} + \frac{\sigma_e^2}{r}$$

$$F_1 : V(\bar{Y}_{3.}) = \frac{V_{F_1}}{rK_3} + \frac{\sigma_e^2}{r}$$

$$RC_{11} : V(\bar{Y}_{4.}) = \frac{V_{RC_{11}}}{rK_4} + \frac{\sigma_e^2}{r}$$

$$RC_{12} : V(\bar{Y}_{5.}) = \frac{V_{RC_{12}}}{rK_5} + \frac{\sigma_e^2}{r}$$

$$F_2 : V(\bar{Y}_{6.}) = \frac{V_{F_2}}{rK_6} + \frac{\sigma_e^2}{r}.$$

As variâncias acima são as que deveriam ser substituídas por suas estimativas na matriz V, e não simplesmente V_{P1} , V_{P2} , V_{F1} , V_{RC11} , V_{RC12} e V_{F2} (conforme a proposta original, sem fazer uso de delineamentos), para uma adequada utilização do método dos quadrados mínimos ponderados. Este detalhe aparentemente não tem sido observado em estudos de herança de uma maneira geral; contudo, é possível que as estimativas não sofram grandes diferenças, por não se construir a matriz V da forma mais adequada, conforme apresentado acima. No presente trabalho, este aspecto também não é levado em consideração, admitindo que os dados não são provenientes de um delineamento experimental.

2.3 Herança Monogênica

No estudo da herança de características contínuas, é freqüentemente interessante ao pesquisador julgar a hipótese da existência de um gene de efeito maior. Quando são observadas plantas de gerações oriundas de linhagens contrastantes (além, eventualmente, das próprias linhagens), então duas metodologias sugeridas na literatura merecem destaque: a proposta por Arias et al. (1994) e a sugerida inicialmente por Souza Sobrinho (1998). Embora tais metodologias apresentem, rigorosamente falando, a mesma base conceitual, possuem particularidades que justificam sua apresentação em separado, o que é feito a seguir.

2.3.1. Metodologia de Arias et al. (1994)

A motivação de Arias et al. (1994) originou-se de estudos de resistência a doenças em soja, nos quais, em virtude de o efeito do ambiente ser pronunciado, o limite entre as classes correspondentes a resistência e susceptibilidade não é claramente definido. Assim, os autores sugerem que classes de resistência sejam definidas (por exemplo, mediante o uso de notas), categorizando as observações das diferentes gerações, conforme representado na Tabela 2, para as gerações P_1 , P_2 , F_1 e F_2 .

TABELA 2 - Distribuição de freqüências absolutas observadas em K classes C_j , nas gerações P_1 , P_2 , F_1 e F_2 . Adaptado de Arias et al. (1994)¹

GERAÇÃO	CLASSE				TOTAIS
	C_1	C_2	...	C_K	
P_1	N_{11}	N_{12}	...	N_{1K}	N_1
P_2	N_{21}	N_{22}	...	N_{2K}	N_2
F_1	N_{31}	N_{32}	...	N_{3K}	N_3
F_2	N_{41}	N_{42}	...	N_{4K}	N_4

¹ N_{ij} é o número de indivíduos da geração i observados na classe j .

A metodologia proposta pelos autores pressupõe que, em presença de um gene de efeito maior, as freqüências em cada classe da geração F_2 deveriam ser função daquelas freqüências das gerações P_1 , F_1 e P_2 , com proporções relativas esperadas 1:2:1, respectivamente. Assim, freqüências absolutas esperadas na classe j da geração F_2 poderiam ser calculadas utilizando a expressão:

$$fe_j = \left[\frac{N_{1j}}{N_1} + 2 \frac{N_{3j}}{N_3} + \frac{N_{2j}}{N_2} \right] \frac{N_4}{4} \quad (1.7).$$

Em posse de tais frequências, uma estatística qui-quadrado poderia ser então calculada:

$$\chi_c^2 = \sum_{j=1}^K \frac{(fe_j - N_{4j})^2}{fe_j} \quad (1.8),$$

com a ressalva de que, se $fe_j < 5$, para algum j , então classes podem ser fundidas, conforme usualmente sugerido em testes de aderência. Arias et al. (1994), no contexto de resistência a doenças, utilizam 6 classes, relativas à porcentagem de área foliar lesionada, as quais foram posteriormente reunidas em apenas duas, correspondentes a resistentes e suscetíveis, utilizando um ponto de truncamento que, segundo os autores, pode variar de estudo para estudo. Sob a hipótese de nulidade (um único gene), a estatística acima tem distribuição aproximada de qui-quadrado com $k-1$ graus de liberdade, sendo k o número de classes efetivamente consideradas na geração F_2 . Deve-se perceber que os dados da geração F_2 não são utilizados para estimar parâmetros de cálculo de frequências esperadas, e é por isso que apenas 1 grau de liberdade é subtraído do número de classes.

Embora não ressaltado pelos autores, evidentemente esta metodologia pode ser facilmente adaptada quando se dispõe de outras gerações segregantes, como retrocruzamentos.

2.3.2. Metodologia Apresentada por Souza Sobrinho (1998)

A metodologia originalmente apresentada por Souza Sobrinho (1998) foi também adaptada por outros autores, como Resende (1999), Gomes (1999), Azevedo (2001) e Freitas et al. (2002), por vezes com pequenas modificações, devido a particularidades de cada trabalho. Assim, optou-se aqui por apresentar a metodologia conforme descrito por Freitas et al. (2002).

Estes últimos autores consideraram dados de gerações P_1 , P_2 , F_1 e F_2 , mas a descrição que segue pode ser facilmente adaptada para contemplar outras gerações. A característica estudada pelos autores era claramente contínua (teor foliar de zingibereno).

Inicialmente, um ponto de truncamento (PT) é estabelecido, de maneira a claramente discriminar os genitores contrastantes. A metodologia leva em conta diferentes valores possíveis de grau médio de dominância (GMD), e admite que a característica sob estudo tenha distribuição normal. É implementada pelas seguintes etapas (adaptado de Freitas et al., 2002):

1. Estimam-se a média e variância das gerações P_1 e P_2 , por meio de estimadores usuais. Utilizando tais estimativas, calculam-se as frequências esperadas, conforme a densidade normal, abaixo e acima do PT nas gerações P_1 e P_2 ;

2. Com cada valor especificado de GMD, estima-se a média da geração F_1 por:

$$\bar{F}_1 = (\bar{P}_1 + \bar{P}_2)/2 + GMD(\bar{P}_1 - \bar{P}_2) \quad (1.9),$$

sendo \bar{P}_1 e \bar{P}_2 as médias das gerações P_1 e P_2 . A variância na geração F_1 é estimada pela variância amostral entre plantas F_1 . A partir da densidade normal com tais parâmetros, frequências esperadas abaixo e acima de PT são também calculadas;

3. Na geração F_2 , as freqüências esperadas abaixo e acima de PT são calculadas a partir das freqüências esperadas obtidas para as gerações P_1 , F_1 e P_2 , com pesos que mantêm as proporções relativas 1:2:1, respectivamente;

4. Freqüências esperadas em todas as gerações são confrontadas com as freqüências observadas mediante uma estatística de qui-quadrado e um teste de aderência é feito;

5. A estatística de qui-quadrado é calculada com diferentes valores de GMD, e os valores menores do que o nível nominal permitem que se infira que a hipótese de herança monogênica deva ser rejeitada.

Os autores ainda mencionam a possibilidade de fusão de gerações, para evitar a ocorrência de freqüências esperada nulas.

É interessante destacar que, no teste de aderência descrito acima, os parâmetros estimados (médias, variâncias e GMD) não são funções diretas das freqüências absolutas observadas nas classes abaixo e acima do PT. Em tais casos, segundo Mood et al. (1974), o número de graus de liberdade da distribuição qui-quadrado sob a hipótese de nulidade é algo entre o valor $K-1$ e $k-1-p$, em que k é o número de classes e p é o número de parâmetros estimados. No presente contexto, obviamente o termo $(K-1)$ que aparece nestes extremos deveria ser multiplicado pelo número de gerações efetivamente envolvidas no teste de aderência.

Por exemplo, no estudo de Freitas et al. (2002), as gerações P_1 , P_2 e F_1 foram fundidas em uma única distribuição de freqüência. Em posse desta distribuição, bem como da geração F_2 , e lembrando que $K = 2$ em cada uma delas, tem-se que o número de graus de liberdade da estatística de teste é um valor ente 0 (uma vez que não pode ser negativo) e 2. Os autores utilizaram o valor intermediário 1.

Como se pode observar, as metodologias de Arias et al. (1994) e de Souza Sobrinho (1998) apresentam a mesma base conceitual, no sentido de que

freqüências esperadas são calculadas nas gerações segregantes utilizando informações das gerações P_1 , P_2 e F_1 , e baseando-se em um comportamento mendeliano de um único gene. As principais diferenças resultam do fato de que a metodologia de Souza Sobrinho (1998) admite distribuição normal e fornece explicitamente um estimador do GMD (dado por aquele valor ao qual corresponde um mínimo qui-quadrado) que faz uso de todas as gerações, e não apenas de P_1 , P_2 e F_1 .

Também é interessante observar que, na metodologia proposta por Souza Sobrinho (1998), por se tratar de um teste de aderência, várias classes podem ser definidas para confrontar freqüências esperadas e observadas, e não apenas duas por geração, correspondentes à divisão abaixo e acima de PT. Ou seja, pontos arbitrários poderiam ser definidos para construir classes. Isso poderia contribuir, eventualmente, para aumentar o poder da metodologia, pelo aumento no número de graus de liberdade.

Outro ponto em comum entre as metodologias é o fato de que a maior parte ou toda a informação necessária para o cálculo das freqüências esperadas é obtida das gerações P_1 , P_2 e F_1 , essencialmente devido aos valores genotípicos dos homozigotos e do heterozigoto estarem sendo estimados utilizando basicamente apenas estas gerações. Seria interessante que a informação sobre tais valores genotípicos, contida em F_2 , bem como outras gerações segregantes, fosse aproveitada no processo de estimação. O enfoque estatístico natural seria reconhecer que as distribuições de probabilidade nestas gerações são misturas de densidades, em geral normais.

2.4. Modelos de Misturas

Aitkin & Wilson (1980) apresentam o processo básico de estimação de parâmetros em misturas de distribuições de normais e consideram, como

exemplos, os casos que envolvem dois componentes da mistura com iguais variâncias, dois componentes com médias iguais e variâncias diferentes, dois componentes com médias e variâncias diferentes e, por fim, envolvendo mais do que dois componentes.

Os autores mostram que o processo de obtenção de estimativas de máxima verossimilhança é o algoritmo EM (Dempster et al., 1977), que alterna passos de esperança e maximização. A título de ilustração, sejam observações independentes e identicamente distribuídas geradas de uma população com função de densidade:

$$f(y) = pf_1(y) + (1 - p)f_2(y) \quad (1.10),$$

em que:

$$f_j(y | \mu_j, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(y - \mu_j)^2 / (2\sigma^2)], \quad j = 1, 2 \quad (1.11).$$

Os estimadores de máxima verossimilhança dos parâmetros μ_1, μ_2, σ^2 e p correspondem às seguintes soluções:

$$\hat{p} = \sum_i \hat{P}(1 | y_i) / n$$

$$\hat{\mu}_j = \sum_i y_i \hat{P}(j | y_i) / \sum_i \hat{P}(j | y_i)$$

$$\hat{\sigma}^2 = \sum_j \sum_i \{(y_i - \hat{\mu}_j)^2 \hat{P}(j | y_i)\} / n, \quad (j=1, 2, e i=1, 2, \dots, n),$$

sendo $\hat{p}(j | y_i)$ a estimativa da probabilidade de que y_i seja uma realização do j -ésimo componente da mistura:

$$\begin{aligned} \hat{P}(1 | y_i) &= \hat{p}\hat{f}_1(y_i) / \{\hat{p}\hat{f}_1(y_i) + (1 - \hat{p})\hat{f}_2(y_i)\} \\ &= 1 - \hat{P}(2 | y_i). \end{aligned} \quad (1.12)$$

Os valores estimados são obtidos pelo método numérico iterativo, o algoritmo EM. Dados $P(1 | y_i)$, a estimação de p, μ e σ^2 corresponde à etapa

de maximização, e a partir destas estimativas, os $P(j|y_i)$ são recalculados (etapa de esperança).

No contexto da genética, modelos de misturas de normais são usados para a estimação de parâmetros relativos a um gene de efeito maior (Lynch & Walsh, 1998; Capítulo 13). Estes autores consideram uma população na qual as frequências genótípicas deste gene seguem o modelo de equilíbrio de Hardy-Weinberg. A natureza de mistura vem do fato de que a cada genótipo do gene de efeito maior corresponde um valor genotípico esperado, em torno do qual ocorre uma variação aleatória. Esta variação se deve à ação de efeitos ambientais somados ou não à ação de poligenes.

Os pesos da mistura neste enfoque correspondem às frequências genótípicas, que, por sua vez, dependem da frequência alélica de algum dos alelos; esta última é, assim, um parâmetro adicional a ser estimado. Os autores consideram ainda diferentes modelos hierárquicos, desde uma situação mais simples (apenas efeitos ambientais) até uma condição em presença de um gene de efeito maior e também de poligenes modificadores.

Wu et al. (2001) estendem o enfoque de detecção de genes maiores para situações de cruzamentos controlados, levando em conta a possibilidade de alelismo múltiplo para o gene de efeito maior, bem como a eventual ocorrência de desequilíbrio de Hardy-Weinberg.

Tanto Lynch & Walsh (1998) como Wu et al. (2001) fazem uso do algoritmo EM para a obtenção das estimativas de máxima verossimilhança dos parâmetros.

Um outro contexto em genética, no qual modelos de misturas são muito utilizados, em associação ao algoritmo EM, é o mapeamento de QTLs (e.g., Zeng, 1994). A natureza da mistura aqui se deve ao fato de que, a cada genótipo quanto a locos marcadores, ocorrem diferentes genótipos quanto a QTLs, devido à recombinação.

2.4.1. Métodos Numéricos Alternativos

Quando a variação (ambiental e eventualmente poligênica) é de mesma natureza dentro de cada componente da mistura, como nos modelos de Lynch & Walsh (1998) e de Wu et al. (2001), a implementação do algoritmo EM é relativamente simples. Por outro lado, quando a amostra apresenta categorias dentro das quais a variação é de natureza particular (por exemplo, ao utilizar gerações P_1 , P_2 , F_1 , F_2 , RC_{11} , RC_{12} , etc), então a obtenção das expressões correspondentes aos passos E e M não é, em absoluto, imediata. Na realidade, obter expressões fechadas, como nos modelos anteriores, não é aparentemente possível, pelo menos não sem consideráveis adaptações.

Assim, nesses casos, pode ser mais apropriado fazer uso de outros algoritmos numéricos para a obtenção das estimativas de máxima verossimilhança. Alguns dos métodos que poderiam ser utilizados para tal são os métodos de Newton-Raphson, Quase Newton e o de Powell (Kennedy & Gentle, 1980).

O método de Newton-Raphson requer a matriz de derivadas segundas da função suporte. Portanto, uma vantagem do método é a de que ele adicionalmente fornece um critério para a estimação dos elementos da matriz de variâncias e covariâncias dos estimadores, uma vez que esta é a inversa da matriz de derivadas segundas da função suporte, multiplicada por (-1).

Assim, fazendo uso da propriedade de normalidade assintótica dos estimadores de máxima verossimilhança, é imediata a construção de intervalos de confiança aproximados para os parâmetros do modelo, ao se utilizar o método de Newton-Raphson.

No entanto, quando há mais de um parâmetro no modelo, é comum que essa matriz, em uma dada iteração, não seja positiva definida. A consequência

disto é a de que, em uma iteração seguinte, a função objetivo (no caso, a função suporte) não se aproxima do ponto de máximo.

Para contornar este problema, os chamados métodos Quase-Newton foram propostos no sentido de substituir esta matriz de derivadas segunda por alguma outra matriz convenientemente construída, de maneira a ser sempre positiva definida em todas as iterações. Assim, os métodos de Quase-Newton exigem apenas as derivadas primeiras da função objetivo. Segundo Kennedy & Gentle (1980), os métodos Quase-Newton já poderiam ser considerados, àquela época, como os mais populares no problema geral de obtenção de pontos ótimos de função de mais de uma variável. Dentre os critérios de construção da matriz positiva definida destaca-se o algoritmo de Davidon-Fletcher-Powell, conforme descrito pelos mesmos autores.

O método de Powell não exige nem mesmo o cálculo das derivadas primeiras, e também foi descrito por Kennedy & Gentle (1980). Sendo p o número de parâmetros do modelo, o método utiliza p iterações para obter um conjunto de direções conjugadas que levaria a um mínimo se a função objetivo fosse absolutamente quadrática. Este processo de p iterações pode ser repetido até a convergência. Por não levar em conta o cálculo de derivadas, o método de direções conjugadas é consideravelmente mais lento.


3 METODOLOGIA

Este capítulo apresenta uma metodologia para modelar e estimar parâmetros relativos à ação de genes de efeito maior e poligenes ao se utilizarem gerações derivadas de linhagens contrastantes. Embora, rigorosamente falando, os princípios aqui envolvidos sejam os mesmos dos modelos de Lynch & Walsh (1998) e de Wu et al. (2001), existem algumas particularidades ao se utilizarem estas gerações mencionadas que justificam uma formalização particular. Dentre essas particularidades, destacam-se a natureza diferenciada da variação fenotípica em cada geração e a dificuldade de implementação do algoritmo EM, como feita em geral.

Aqui são consideradas as gerações P_1 , P_2 , F_1 , RC_{11} , RC_{12} e F_2 , mas o enfoque pode ser estendido para outras gerações derivadas de linhagens contrastantes. Além de apresentar modelos probabilísticos para aquelas gerações, aqui é considerada também a estimação dos parâmetros pelo método da máxima verossimilhança.

A função de verossimilhança foi constituída considerando uma característica que pode apresentar um gene de efeito maior, sendo que há variação entre indivíduos de um mesmo genótipo devido à ação de efeitos ambientais e/ou de genes de efeitos menores. -

Admitiu-se distribuição de probabilidade normal, a cada genótipo, com média e variância específicas, sendo que às populações F_2 , RC_{11} e RC_{12} correspondem misturas de distribuições normais, com pesos atribuídos de acordo com a frequência esperada em cada geração, dos diferentes genótipos.



3.1 Construção do Modelo Genético

Na construção do modelo genético, considerou-se como o modelo mais geral aquele que apresenta a existência de gene de efeito maior mais polígenes com efeitos aditivos e de dominância, e variâncias ambientais (σ^2) iguais em todas as gerações. Admitiram-se ainda genes independentes (tanto polígenes como de efeito maior).

No caso da existência de gene de efeito maior, os valores genotípicos dos dois homocigotos e do heterocigoto são representados, respectivamente, por $\mu - A$, $\mu + A$ e $\mu + D$, sendo μ uma constante de referência, A o efeito aditivo do gene de efeito maior e D o efeito de dominância do gene de efeito maior. Com a existência de polígenes modificadores, têm-se os componentes poligênicos de média e variância conforme Tabela 3, sendo $[a]$ o afastamento da média dos pais homocigotos da linhagem com maior expressão do caráter ou componente poligênico aditivo, $[d]$ o afastamento da média dos pais para os heterocigotos ou componente poligênico de dominância, V_A a variância aditiva, V_D a variância atribuída aos desvios de dominância dos efeitos poligênicos e S_{AD} o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância (Mather & Jinks, 1984).

A derivação dos termos apresentados na Tabela 3 é dada como segue. Se $\bar{P}_1, \bar{P}_2, \bar{F}_1, \bar{RC}_{11}, \bar{RC}_{12}$ e \bar{F}_2 representam as esperanças de cada uma dessas gerações, considerando apenas a existência de polígenes, então, segundo Mather & Jinks (1984):

$$\bar{RC}_{11} = \frac{1}{2}(\bar{F}_1 + \bar{P}_1),$$

$$\bar{RC}_{12} = \frac{1}{2}(\bar{F}_1 + \bar{P}_2)$$

$$e \quad \bar{F}_2 = \frac{1}{4}(2\bar{F}_1 + \bar{P}_1 + \bar{P}_2).$$

Sendo os componentes poligênicos de médias para as gerações P_1 , P_2 e F_1 , respectivamente $-[a]$, $[a]$ e $[d]$, então os componentes poligênicos de média para os retrocruzamentos e a geração F_2 , segundo Mather & Jinks (1984), são dados por:

$$\overline{RC}_{11} = \frac{1}{2}(-[a] + [d]) = -\frac{1}{2}[a] + \frac{1}{2}[d],$$

$$\overline{RC}_{12} = \frac{1}{2}([a] + [d]) = \frac{1}{2}[a] + \frac{1}{2}[d]$$

$$e \quad \bar{F}_2 = \frac{1}{4}(2[d] - [a] + [a]) = \frac{1}{2}[d],$$

conforme a Tabela 3.

TABELA 3 - Gerações envolvidas e seus respectivos componentes poligênicos da média e da variância

GERAÇÃO	Componente poligênico da	
	MÉDIA	Componente poligênico da VARIÂNCIA
P_1	$-[a]$	—
P_2	$[a]$	—
F_1	$[d]$	—
RC_{11}	$-\frac{1}{2}[a] + \frac{1}{2}[d]$	$\frac{1}{2}V_A + V_D - S_{AD}$
RC_{12}	$\frac{1}{2}[a] + \frac{1}{2}[d]$	$\frac{1}{2}V_A + V_D + S_{AD}$
F_2	$\frac{1}{2}[d]$	$V_A + V_D$

Os componentes poligênicos da variância para as gerações P_1 , P_2 e F_1 não existem, pois toda variação tem que ser exclusivamente não-herdável, uma

vez que todos os indivíduos em cada geração dessa possuem o mesmo genótipo. Os componentes poligênicos das variâncias $V_{RC_{11}}$, $V_{RC_{12}}$ e V_{F_2} , respectivamente para RC_{11} , RC_{12} e F_2 , segundo Mather & Jinks (1984), são dados por:

$$V_{RC_{11}} = \frac{1}{4}S(a^2) + \frac{1}{4}S(d^2) - \frac{1}{2}S(ad) + \sigma^2,$$

$$V_{RC_{12}} = \frac{1}{4}S(a^2) + \frac{1}{4}S(d^2) + \frac{1}{2}S(ad) + \sigma^2$$

e
$$V_{F_2} = \frac{1}{2}S(a^2) + \frac{1}{4}S(d^2) + \sigma^2,$$

sendo σ^2 um componente não-herdável, $S(a^2)$ a soma dos quadrados das medidas dos afastamentos dos homozigotos da média dos pais (efeitos aditivos) e $S(d^2)$ a soma dos quadrados das medidas dos afastamentos dos heterozigotos da média dos pais (efeitos de dominância), e $S(ad)$ a soma dos produtos dos efeitos aditivos pelos de dominância.

Sendo $V_A = \frac{1}{2}S(a^2)$, $V_D = \frac{1}{4}S(d^2)$ e $S_{AD} = \frac{1}{2}S(ad)$, substituindo

nas expressões de componentes poligênicos de variância, tem-se:

$$V_{RC_{11}} = \frac{1}{2}V_A + V_D - S_{AD} + \sigma^2,$$

$$V_{RC_{12}} = \frac{1}{2}V_A + V_D + S_{AD} + \sigma^2$$

e
$$V_{F_2} = V_A + V_D + \sigma^2,$$

conforme foi apresentado na Tabela 3.

Seja agora uma amostra aleatória com elementos Y_{ij} , que representa o valor observado do indivíduo i ($i = 1, 2, \dots, n_j$) da geração j ($j = 1, 2, \dots, 6$, que corresponde, respectivamente às gerações P_1 , P_2 , F_1 , RC_{11} , RC_{12} e F_2).

Admitindo distribuição normal, as funções de densidade de probabilidade são assim estruturadas:

$$P_1: f_1(y_{i1}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_{i1} - \mu + [a] + A)^2}{2\sigma^2}\right\},$$

$$P_2: f_2(y_{i2}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_{i2} - \mu - [a] - A)^2}{2\sigma^2}\right\},$$

$$F_1: f_3(y_{i3}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_{i3} - \mu - [d] - D)^2}{2\sigma^2}\right\},$$

$$RC_{11}: f_4(y_{i4}) = \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D - S_{AD}}} \exp\left\{-\frac{(y_{i4} - \mu + \frac{[a]}{2} - \frac{[d]}{2} + A)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D - S_{AD})}\right\} +$$

$$+ \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D - S_{AD}}} \exp\left\{-\frac{(y_{i4} - \mu + \frac{[a]}{2} - \frac{[d]}{2} - D)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D - S_{AD})}\right\},$$

$$RC_{12}: f_5(y_{i5}) = \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D + S_{AD}}} \exp\left\{-\frac{(y_{i5} - \mu - \frac{[a]}{2} - \frac{[d]}{2} - A)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D + S_{AD})}\right\} +$$

$$+ \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D + S_{AD}}} \exp\left\{-\frac{(y_{i5} - \mu - \frac{[a]}{2} - \frac{[d]}{2} - D)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D + S_{AD})}\right\},$$

$$F_2: f_6(y_{i6}) = \frac{1}{4} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + V_A + V_D}} \exp\left\{-\frac{(y_{i6} - \mu - \frac{[d]}{2} + A)^2}{2(\sigma^2 + V_A + V_D)}\right\} +$$

$$+ \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + V_A + V_D}} \exp\left\{-\frac{(y_{i6} - \mu - \frac{[d]}{2} - D)^2}{2(\sigma^2 + V_A + V_D)}\right\} +$$

$$+ \frac{1}{4} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + V_A + V_D}} \exp \left\{ -\frac{(y_{i6} - \mu - \frac{[d]}{2} - A)^2}{2(\sigma^2 + V_A + V_D)} \right\}.$$

Note que as funções de densidade para RC₁₁ e RC₁₂ são constituídas pela mistura de duas distribuições normais e F₂ por uma mistura de três distribuições normais, sendo que, em cada componente da mistura, os componentes de média e de variância dos poligenes não mudam, mudando apenas os efeitos do gene de efeito maior. Ou seja, é a existência deste que determina a ocorrência destas misturas.

3.2. Estimação

A partir do modelo descrito no item anterior, os parâmetros podem ser obtidos pelo método de máxima verossimilhança. A função de verossimilhança (L) é assim definida:

$$L = \prod_{j=1}^6 \prod_{i=1}^{n_j} f(y_{ij}) \quad (1.13).$$

Aplicando o logaritmo na função L, obtém-se a denominada função Suporte (S),

$$S = \sum_{j=1}^6 \sum_{i=1}^{n_j} \ln f(y_{ij}) \quad (1.14).$$

De outra forma, tem-se:

$$S = \sum_{i=1}^{n_1} \ln \phi(y_{i1}; \mu - [a] - A, \sigma^2) + \sum_{i=1}^{n_2} \ln \phi(y_{i2}; \mu + [a] + A, \sigma^2) + \\ \sum_{i=1}^{n_3} \ln \phi(y_{i3}; \mu + [d] + D, \sigma^2) + \sum_{i=1}^{n_4} \ln \left\{ \frac{1}{2} \phi(y_{i4}; \mu - \frac{[a]}{2} + \frac{[d]}{2} - A, \sigma^2 + \frac{V_A}{2} + V_D - S_{AD}) + \right. \\ \left. + \frac{1}{2} \phi(y_{i4}; \mu - \frac{[a]}{2} + \frac{[d]}{2} - A, \sigma^2 + \frac{V_A}{2} + V_D - S_{AD}) + \right\}$$

$$\begin{aligned}
& + \frac{1}{2} \phi(y_{i4}; \mu - \frac{[a]}{2} + \frac{[d]}{2} + D, \sigma^2 + \frac{V_A}{2} + V_D - S_{AD}) \} + \\
& \sum_{i=1}^{n_5} \ln \{ \frac{1}{2} \phi(y_{i5}; \mu + \frac{[a]}{2} + \frac{[d]}{2} + A, \sigma^2 + \frac{V_A}{2} + V_D + S_{AD}) \} + \\
& + \frac{1}{2} \phi(y_{i5}; \mu + \frac{[a]}{2} + \frac{[d]}{2} + D, \sigma^2 + \frac{V_A}{2} + V_D + S_{AD}) \} + \\
& + \sum_{i=1}^{n_6} \ln \{ \frac{1}{4} \phi(y_{i6}; \mu + \frac{[d]}{2} - A, \sigma^2 + V_A + V_D) + \frac{1}{2} \phi(y_{i6}; \mu + \frac{[d]}{2} + D, \sigma^2 + V_A + V_D) + \\
& \quad \frac{1}{4} \phi(y_{i6}; \mu + \frac{[d]}{2} + A, \sigma^2 + V_A + V_D) \},
\end{aligned}$$

na qual $\phi(y; M, Var)$ representa a densidade normal com média M e variância Var.

Derivando esta função suporte (S) em relação a cada um dos parâmetros do modelo e em seguida igualando estas expressões a zero, obtêm-se as equações de máxima verossimilhança. Além disso, para a obtenção de uma estimativa da matriz de informação de Fisher, e conseqüentemente das variâncias e covariâncias dos estimadores, é conveniente a obtenção das derivadas segundas da função suporte. Eventualmente, estas poderiam também ser utilizadas na implementação do método de Newton-Raphson.

Assim, a seguir são apresentadas as expressões gerais das derivadas parciais primeiras e segundas, em relação a componentes de médias e de variâncias do modelo.

3.2.1 Derivada Parcial Primeira em Relação a um Componente Qualquer de Médias

Seja o parâmetro θ um componente qualquer de média, g uma função de θ , e V a variância em uma densidade normal. A derivada primeira da função de densidade de probabilidade ϕ , em relação a θ , é dada pela expressão:

$$\frac{\partial \ln \phi[y; g_{(j)l}(\theta), V_j]}{\partial \theta} = \frac{1}{\phi[y; g_{(j)l}(\theta), V_j]} \phi[y; g_{(j)l}(\theta), V_j] \frac{\partial}{\partial \theta} \left[-\frac{(y - g_{(j)l}(\theta))^2}{2V_j} \right] =$$

$$-\frac{1}{V_j} [y - g_{(j)l}(\theta)] (-1) g'_{(j)l}(\theta) = \frac{[y - g_{(j)l}(\theta)] g'_{(j)l}(\theta)}{V_j} \quad (1.15),$$

em que $g'_{(j)l}(\theta) = \frac{\partial g_{(j)l}(\theta)}{\partial \theta}$.

Uma geração qualquer j possui uma distribuição de probabilidade que é uma mistura de K_j densidades normais, lembrando que em algumas gerações K_j pode ser igual a 1. Assim, de uma forma geral, em uma geração qualquer, cada componente l da mistura ($l = 1, 2, \dots, K_j$) terá uma média $\mu_{(j)l}$, que pode ser expressa como uma função linear de θ , na forma $g_{(j)l}(\theta)$. Assim:

$$\frac{\partial}{\partial \theta} \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), V_j] =$$

$$\frac{1}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), V_j]} \sum_{l=1}^{k_j} \frac{w_{(j)l} \phi[y; g_{(j)l}(\theta), V_j] [y - g_{(j)l}(\theta)] g'_{(j)l}(\theta)}{V_j} \quad (1.16),$$

em que $w_{(j)l}$ corresponde aos pesos (frequências) associados ao l -ésimo componente da mistura da j -ésima geração, e V_j a variância dentro de cada componente da geração j .

3.2.2 Derivada Parcial Primeira em Relação a um Componente Qualquer da Variância

Tomando-se τ como um componente qualquer da variância na função de densidade de probabilidade normal ϕ , tem-se, então, que a variância é uma função de τ , ou seja, $h_j(\tau)$. Assim, na geração j :

$$\begin{aligned}
 \frac{\partial}{\partial \tau} \sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau)] &= \sum_{l=1}^{k_j} w_{(j)l} \frac{\partial \phi[y; \mu_{(j)l}, h_j(\tau)]}{\partial \tau} = \\
 &= \sum_{l=1}^{k_j} w_{(j)l} \left\{ \phi[y; \mu_{(j)l}, h_j(\tau)] \frac{\partial}{\partial \tau} \left[-\frac{(y - \mu_{(j)l})^2}{2h_j(\tau)} \right] + \right. \\
 &\quad \left. + \left(-\frac{1}{2}\right) \frac{1}{h_j(\tau)} \frac{1}{\sqrt{2\pi h_j(\tau)}} \exp \left[-\frac{(y - \mu_{(j)l})^2}{2h_j(\tau)} \right] h_j'(\tau) \right\} = \\
 \sum_{l=1}^{k_j} w_{(j)l} \left\{ \phi[y; \mu_{(j)l}, h_j(\tau)] \frac{(y - \mu_{(j)l})^2}{2[h_j(\tau)]^2} h_j'(\tau) - \frac{1}{2h_j(\tau)} \phi[y; \mu_{(j)l}, h_j(\tau)] h_j'(\tau) \right\} &= \\
 \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau)] h_j'(\tau)}{2h_j(\tau)} \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau)} - 1 \right] & \quad (1.17),
 \end{aligned}$$

em que $h_j'(\tau) = \frac{\partial h_j(\tau)}{\partial \tau}$.

Dessa forma:

$$\begin{aligned}
 \frac{\partial}{\partial \tau} \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau)] &= \\
 \frac{\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau)] h_j'(\tau)}{2h_j(\tau)} \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau)} - 1 \right]}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau)]} & \quad (1.18).
 \end{aligned}$$

3.2.3 Derivada Parcial Segunda em Relação a dois Componentes Quaisquer de Média

Considerando que a média da função de densidade de probabilidade ϕ é função de dois componentes θ_1 e θ_2 quaisquer de média, no componente de mistura l da geração j , representada por $g_{(j)l}(\theta_1, \theta_2)$, então:

$$\frac{\partial}{\partial \theta_1} \ln \sum_{j=1}^k w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] = \frac{\sum_{j=1}^k w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] [y - g_{(j)l}(\theta_1, \theta_2)] g'_{(j)l}(\theta_1)}{V_j \sum_{j=1}^k w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]},$$

conforme resultado anterior.

Na expressão acima, como a derivada de $g_{(j)l}(\theta_1, \theta_2)$ em relação a θ_1 não depende de θ_2 , daí decorre a notação $g'_{(j)l}(\theta_1)$.

Para a derivada segunda da expressão anterior em relação a θ_2 , tem-se:

$$\frac{\partial^2 \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]}{\partial \theta_1 \partial \theta_2} = \frac{\partial}{\partial \theta_2} \left\{ \frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] [y - g_{(j)l}(\theta_1, \theta_2)] g'_{(j)l}(\theta_1)}{V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]} \right\}.$$

Denominando os termos envolvidos na derivação em relação a θ_2 por u , z e v , isto é,

$$u = w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]$$

$$z = [y - g_{(j)l}(\theta_1, \theta_2)] g'_{(j)l}(\theta_1)$$

$$e \quad v = V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j];$$

assim, pode-se representar a derivada segunda, de forma geral, por:

$$\frac{\partial^2 \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]}{\partial \theta_1 \partial \theta_2} = \left[\frac{\sum_{l=1}^{k_j} u \cdot z}{v} \right]' =$$

$$\frac{v \cdot \left(\sum_{l=1}^{k_j} u \cdot z \right)' - v' \cdot \sum_{l=1}^{k_j} u \cdot z}{v^2} = \frac{v \cdot \sum_{l=1}^{k_j} u \cdot z' + v \cdot \sum_{l=1}^{k_j} z u' - v' \cdot \sum_{l=1}^{k_j} u \cdot z}{v^2}$$

Derivando u , z e v em relação a θ_2 , obtêm-se:

$$u' = \frac{\partial w_{(j)l} \phi[y; g_{(j)l}(\theta^1, \theta^2), V_j]}{\partial \theta_2} =$$

$$w_{(j)l} \phi[y; g_{(j)l}(\theta^1, \theta^2), V_j] \frac{[y - g_{(j)l}(\theta_1, \theta_2)]}{V_j} g'_{(j)l}(\theta_2),$$

$$z' = \frac{\partial [y - g_{(j)l}(\theta_1, \theta_2)] g'_{(j)l}(\theta^1)}{\partial \theta_2} = -g'_{(j)l}(\theta_2) g'_{(j)l}(\theta_1) \quad e$$

$$v' = \frac{\partial V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]}{\partial \theta_2} = V_j \sum_{l=1}^{k_j} w_{(j)l} \frac{\partial}{\partial \theta_2} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] =$$

$$V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] \frac{[y - g_{(j)l}(\theta_1, \theta_2)]}{V_j} g'_{(j)l}(\theta_2) =$$

$$\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] [y - g_{(j)l}(\theta_1, \theta_2)] g'_{(j)l}(\theta_2),$$

que substituídas na expressão principal, resultam em:

$$\begin{aligned}
& \frac{\partial^2 \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]}{\partial \theta_1 \partial \theta_2} = \\
& \frac{V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] [-g'_{(j)l}(\theta_2)] g'_{(j)l}(\theta_1)}{[V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]]^2} + \\
& \frac{V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] \sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta_1, \theta_2)]^2 \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] \frac{g'_{(j)l}(\theta_2)}{V_j} g'_{(j)l}(\theta_1)}{[V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]]^2} + \\
& - \frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] [y - g_{(j)l}(\theta_1, \theta_2)] g'_{(j)l}(\theta_2)}{[V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j]]^2} \times \\
& \times \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta_1, \theta_2), V_j] [y - g_{(j)l}(\theta_1, \theta_2)] g'_{(j)l}(\theta_1) \quad (1.19).
\end{aligned}$$

3.2.4 Derivada Parcial Segunda em Relação a um Componente Qualquer de Média e um Componente Qualquer de Variância

Esta derivada segunda corresponde a:

$$\frac{\partial^2 \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)]}{\partial \theta \partial \tau} =$$

$$= \frac{\partial \sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta)] g'_{(j)l}(\theta) \phi[y; g_{(j)l}(\theta), h_j(\tau)]}{\partial \tau \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau)}$$

A derivada em relação ao componente de variância τ é do tipo:

$$\frac{v \sum u' - v' \sum u}{v^2}, \text{ em que } u = \sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta)] g'_{(j)l}(\theta) \phi[y; g_{(j)l}(\theta), h_j(\tau)]$$

$$\text{e } v = \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau); \text{ portanto:}$$

$$u' = \frac{\partial}{\partial \tau} \sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta)] g'_{(j)l}(\theta) \phi[y; g_{(j)l}(\theta), h_j(\tau)] =$$

$$= \sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta)] g'_{(j)l}(\theta) \frac{\phi[y; g_{(j)l}(\theta), h_j(\tau)] h'_j(\tau)}{2h_j(\tau)} \left[\frac{[y - g_{(j)l}(\theta)]^2}{h_j(\tau)} - 1 \right]$$

$$v' = \frac{\partial}{\partial \tau} \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau) = \sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h'_j(\tau) +$$

$$\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; g_{(j)l}(\theta), h_j(\tau)] h'_j(\tau)}{2h_j(\tau)} \left[\frac{[y - g_{(j)l}(\theta)]^2}{h_j(\tau)} - 1 \right] h_j(\tau).$$

Assim:

$$\frac{\partial^2 \ln \sum_{j=1}^k w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)]}{\partial \theta \partial \tau} =$$

$$\frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau) \sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta)] g'_{(j)l}(\theta)}{[\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau)]} \times$$

$$\begin{aligned}
& \times \frac{\frac{\phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j'(\tau)}{2h_j(\tau)} \left[\frac{[y - g_{(j)l}(\theta)]^2}{h_j(\tau)} - 1 \right]}{\left[\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau) \right]} + \\
& \frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j'(\tau) \sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta)] g_{(j)l}'(\theta) \phi[y; g_{(j)l}(\theta), h_j(\tau)]}{\left[\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau) \right]^2} + \\
& + \frac{\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j'(\tau)}{2} \left[\frac{[y - g_{(j)l}(\theta)]^2}{h_j(\tau)} - 1 \right]}{\left[\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau) \right]} \times \\
& \times \frac{\sum_{l=1}^{k_j} w_{(j)l} [y - g_{(j)l}(\theta)] g_{(j)l}'(\theta) \phi[y; g_{(j)l}(\theta), h_j(\tau)]}{\left[\sum_{l=1}^{k_j} w_{(j)l} \phi[y; g_{(j)l}(\theta), h_j(\tau)] h_j(\tau) \right]} \quad (1.20).
\end{aligned}$$

3.2.5 Derivada Parcial Segunda Relativa a dois Componentes Quaisquer de Variância

Considerando que a variância na função de densidade de probabilidade ϕ é função de dois componentes de variância quaisquer, τ_1 e τ_2 , ou seja, $h_j(\tau_1, \tau_2)$, então:

$$\frac{\partial^2 \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]}{\partial \tau_1 \partial \tau_2} =$$

$$= \frac{\partial}{\partial \tau_2} \frac{\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1) \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 1 \right]}{2h_j(\tau_1, \tau_2)}}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]}$$

A notação $h_j'(\tau_1)$ deve-se ao fato de a derivada de $h_j(\tau_1, \tau_2)$ em relação a τ_1 não depender de τ_2 .

A derivada segunda em relação a τ_2 é do tipo $\left(\frac{u}{v}\right)' = \frac{vu' - uv'}{v^2} =$

$$\frac{u'}{v} - \frac{uv'}{v^2}.$$

Denominando:

$$u = \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1) \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 1 \right]}{2h_j(\tau_1, \tau_2)} =$$

$$\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1) (y - \mu_{(j)l})^2}{2[h_j(\tau_1, \tau_2)]^2} - \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1)}{2h_j(\tau_1, \tau_2)}$$

e $v = \sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]$, tem-se:

a) Derivada da expressão u :

$$u' = \frac{\partial}{\partial \tau_2} \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1) (y - \mu_{(j)l})^2}{2[h_j(\tau_1, \tau_2)]^2} +$$

$$- \frac{\partial}{\partial \tau_2} \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1)}{2h_j(\tau_1, \tau_2)},$$

em que

$$\frac{\partial}{\partial \tau_2} \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1) (y - \mu_{(j)l})^2}{2[h_j(\tau_1, \tau_2)]^2} =$$

$$\begin{aligned}
&= \sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) (y - \mu_{(j)l})^2 \left[\frac{2[h_j(\tau_1, \tau_2)]^2 \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2)}{2h_j(\tau_1, \tau_2)} \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 1 \right]}{4[h(\tau_1, \tau_2)]^4} + \right. \\
&\quad \left. \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] 4h_j(\tau_1, \tau_2) h_j'(\tau_2)}{4[h_j(\tau_1, \tau_2)]^4} \right] = \\
&= \frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) (y - \mu_{(j)l})^2 \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2) [(y - \mu_{(j)l})^2 - 5h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^4}
\end{aligned}$$

e

$$\begin{aligned}
&\frac{\partial}{\partial \tau_2} \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1)}{2h_j(\tau_1, \tau_2)} = \\
&\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) \left[\frac{2h_j(\tau_1, \tau_2) \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2)}{2h_j(\tau_1, \tau_2)} \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 1 \right]}{4[h_j(\tau_1, \tau_2)]^2} + \right. \\
&\quad \left. - \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] 2h_j'(\tau_2)}{4[h_j(\tau_1, \tau_2)]^2} \right] = \\
&\frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) h_j'(\tau_2) \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 3 \right]}{4[h_j(\tau_1, \tau_2)]^2} = \\
&= \frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) h_j'(\tau_2) \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] [(y - \mu_{(j)l})^2 - 3h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^3}
\end{aligned}$$

Assim:

$$u' = \frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) (y - \mu_{(j)l})^2 \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2) [(y - \mu_{(j)l})^2 - 5h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^4} +$$

$$\frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) h_j'(\tau_2) \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] [(y - \mu_{(j)l})^2 - 3h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^3}.$$

b) Derivada da expressão v :

$$v' = \frac{\partial}{\partial \tau_2} \sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] =$$

$$\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2)}{2h_j(\tau_1, \tau_2)} \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 1 \right].$$

Assim, a derivada segunda em relação a τ_2 resulta na expressão:

$$\frac{\partial^2 \ln \sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]}{\partial \tau_1 \partial \tau_2} =$$

$$\frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) (y - \mu_{(j)l})^2 \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2) [(y - \mu_{(j)l})^2 - 5h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^4} +$$

$$\frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^3} +$$

$$\frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) h_j'(\tau_2) \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] [(y - \mu_{(j)l})^2 - 3h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^3} +$$

$$\frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]}{4[h_j(\tau_1, \tau_2)]^2}.$$

$$\begin{aligned}
& \frac{\sum_{l=1}^{k_j} w_{(j)l} h_j'(\tau_1) (y - \mu_{(j)l})^2 \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]}{2[h_j(\tau_1, \tau_2)]^2} \times \\
& \frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]} \times \\
& \times \frac{\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2)}{2h_j(\tau_1, \tau_2)} \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 1 \right]}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]} + \\
& + \frac{\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_1)}{2h(\tau_1, \tau_2)}}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]} \times \\
& \times \frac{\sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)] h_j'(\tau_2)}{2h_j(\tau_1, \tau_2)} \left[\frac{(y - \mu_{(j)l})^2}{h_j(\tau_1, \tau_2)} - 1 \right]}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y; \mu_{(j)l}, h_j(\tau_1, \tau_2)]} \quad (1.21).
\end{aligned}$$

3.2.6 Função Escore do Modelo Genético

A função escore do modelo genético descrito em 3.1 é a derivada da função suporte (S) em relação aos parâmetros de interesse, e pode ser assim definida:

a) Para um componente de média θ em geral, a função escore é dada por:

$$\frac{\partial S}{\partial \theta} = \sum_{j=1}^6 \sum_{i=1}^{n_j} \frac{\sum_{l=1}^{k_j} w_{(j)l} \phi[y_{ij}; g_{(j)l}(\theta), V_j] [y_{ij} - g_{(j)l}] g_{(j)l}'(\theta)}{V_j \sum_{l=1}^{k_j} w_{(j)l} \phi[y_{ij}; g_{(j)l}(\theta), V_j]} \quad (1.22).$$

b) Em relação a um componente de variância τ , de maneira geral, a derivada da função suporte (S) se traduz na expressão:

$$\frac{\partial S}{\partial \tau} = \frac{\sum_{j=1}^6 \sum_{i=1}^{n_j} \sum_{l=1}^{k_j} w_{(j)l} \frac{\phi[y_{ij}; \mu_{(j)l}, h_j(\tau)] h_j'(\tau)}{2h_j(\tau)} \left[\frac{(y_{ij} - \mu_{(j)l})^2}{h_j(\tau)} - 1 \right]}{\sum_{l=1}^{k_j} w_{(j)l} \phi[y_{ij}; \mu_{(j)l}, h_j(\tau)]} \quad (1.23).$$

TABELA 4 - Pesos ($w_{j(l)}$) da densidade normal l ($l= 1, 2, 3$), dentro da geração j ($j=1, 2, \dots, 6$)

Geração	j	$w_{j(1)}$	$w_{j(2)}$	$w_{j(3)}$
P ₁	1	1	0	0
P ₂	2	0	1	0
F ₁	3	0	0	1
RC ₁₁	4	1/2	0	1/2
RC ₁₂	5	0	1/2	1/2
F ₂	6	1/4	1/4	1/2

Das expressões (funções escores) oriundas da derivação da função S em relação a cada parâmetro do modelo, igualadas a zero, obtêm-se as equações de máxima verossimilhança.

3.3 Sub-Modelos

A partir do modelo genético geral, descrito no item 3.1, submodelos contendo menos parâmetros podem ser construídos, os quais são apresentados na Tabela 5. Esta Tabela apresenta o tipo de herança, a existência de gene de efeito maior e/ou poligenes, com efeitos aditivos de dominância e os parâmetros

envolvidos em cada modelo, sendo σ^2 a variância ambiental e os demais parâmetros, como definidos anteriormente.

TABELA 5 - Modelos genéticos e seus respectivos parâmetros.

MODELO	Herança	Gene Maior	Polígenes	PARÂMETROS
1	Maior+Polígenes	dom. e adit.	dom. e adit.	$\mu, A, D, [a], [d], V_A, V_D, S_{AD}, \sigma^2$
2	Maior+Polígenes	dom. e adit.	adit.	$\mu, A, D, [a], V_A, \sigma^2$
3	Maior+Polígenes	adit.	dom. e adit.	$\mu, A, [a], [d], V_A, V_D, S_{AD}, \sigma^2$
4	Maior+Polígenes	adit.	adit.	$\mu, A, [a], V_A, \sigma^2$
5	Só Polígenes	—	dom. e adit.	$\mu, [a], [d], V_A, V_D, S_{AD}, \sigma^2$
6	Só Polígenes	—	adit.	$\mu, [a], V_A, \sigma^2$
7	Só Gene Maior	dom. e adit.	—	μ, A, D, σ^2
8	Só Gene Maior	adit.	—	μ, A, σ^2
9	Nenhum	—	—	μ, σ^2

Dependendo da amostra disponível (nem sempre se dispõe de todas seis gerações), diferentes modelos podem ser ajustados, os quais estão representados na Tabela 6.

TABELA 6 - Possíveis amostras e respectivos modelos mais completos possíveis de serem ajustados, número de diferentes médias (para GM e Pol.) e diferentes variâncias (para GM)

Amostra	n° pops. com μ 's difer. p/ GM	n° pops. com μ 's difer. p/ Pol.	n° pops. com σ^2 's difer. p/ GM	Modelo mais Completo
P ₁	1	1	1	9
P ₂	1	1	1	9
F ₁	1	1	1	9
RC ₁₁	2	1	1	8
RC ₁₂	2	1	1	8
F ₂	3	1	1	7
P ₁ , P ₂	2	2	1	8
P ₁ , F ₁	2	2	1	8
P ₁ , F ₂	3	2	2	2
P ₁ , RC ₁₁	2	2	2	4
P ₁ , RC ₁₂	3	2	2	2
P ₂ , F ₁	2	2	1	8
P ₂ , F ₂	3	2	2	2
P ₂ , RC ₁₁	3	2	2	2
P ₂ , RC ₁₂	2	2	2	4
F ₁ , RC ₁₁	2	2	2	4
F ₁ , F ₂	3	2	2	7
F ₁ , RC ₁₂	2	2	2	4
F ₂ , RC ₁₁	3	2	2	2
F ₂ , RC ₁₂	3	2	2	2
RC ₁₁ , RC ₁₁	3	2	2	2
P ₁ , P ₂ , F ₂	3	3	2	2
P ₁ , P ₂ , F ₁	3	3	1	7
P ₁ , P ₂ , RC ₁₁	3	3	2	2

“... continua ...”

“TABELA 6, Cont.”

Amostra	n° pops. com μ 's difer. p/ GM	n° pops. com μ 's difer. p/ Pol.	n° pops. com σ^2 's difer. p/ GM	Modelo mais Completo
P ₁ , P ₂ , RC ₁₂	3	3	2	2
P ₁ , F ₁ , RC ₁₁	2	3	2	4
P ₁ , F ₁ , F ₂	3	3	2	2
P ₁ , F ₁ , RC ₁₂	3	3	2	2
P ₂ , F ₁ , F ₂	3	3	2	2
P ₂ , F ₁ , RC ₁₁	3	3	2	2
P ₂ , F ₁ , RC ₁₂	2	3	2	4
P ₂ , F ₂ , RC ₁₂	3	3	3	2
F ₁ , F ₂ , RC ₁₁	3	3	3	2
F ₁ , F ₂ , RC ₁₂	3	3	3	2
F ₁ , RC ₁₁ , RC ₁₂	3	3	3	2
F ₂ , RC ₁₁ , RC ₁₂	3	3	3	2
P ₁ , F ₂ , RC ₁₁	3	3	3	2
P ₁ , RC ₁₁ , RC ₁₂	3	3	3	2
P ₁ , F ₂ , RC ₁₂	3	3	3	2
P ₂ , F ₂ , RC ₁₁	3	3	3	2
P ₂ , RC ₁₁ , RC ₁₂	3	3	3	2
P ₁ , F ₁ , F ₂ , RC ₁₂	3	4	3	2
P ₁ , P ₂ , F ₁ , RC ₁₁	3	4	2	2
P ₁ , P ₂ , F ₁ , F ₂	3	4	2	2
P ₁ , P ₂ , F ₂ , RC ₁₁	3	4	3	2
P ₁ , P ₂ , F ₁ , RC ₁₂	3	4	2	2
P ₁ , P ₂ , RC ₁₁ , RC ₁₂	3	4	2	2
P ₁ , F ₁ , F ₂ , RC ₁₂	3	4	3	2
P ₁ , P ₂ , F ₂ , RC ₁₂	3	4	3	2
P ₁ , F ₁ , RC ₁₁ , RC ₁₂	3	4	3	2
P ₂ , F ₁ , F ₂ , RC ₁₂	3	4	3	2

“... continua ...”

“TABELA 6, Cont.”

Amostra	n° pops. com μ 's difer. p/ GM	n° pops. com μ 's difer. p/ Pol.	n° pops. com σ^2 's difer. p/ GM	Modelo mais Completo
P ₂ , F ₁ , RC ₁₁ , RC ₁₂	3	4	3	2
P ₂ , F ₂ , RC ₁₁ , RC ₁₂	3	4	4	1
P ₁ , F ₁ , F ₂ , RC ₁₂	3	4	3	2
P ₂ , F ₁ , RC ₁₁ , RC ₁₂	3	4	3	2
F ₁ , F ₂ , RC ₁₁ , RC ₁₂	3	4	4	1
P ₁ , P ₂ , F ₁ , F ₂ , RC ₁₁	3	5	3	2
P ₁ , P ₂ , F ₁ , F ₂ , RC ₁₂	3	5	3	2
P ₁ , P ₂ , F ₁ , RC ₁₁ , RC ₁₂	3	5	3	2
P ₁ , P ₂ , F ₂ , RC ₁₁ , RC ₁₂	3	5	4	1
P ₁ , F ₁ , F ₂ , RC ₁₁ , RC ₁₂	3	5	4	1
P ₂ , F ₁ , F ₂ , RC ₁₁ , RC ₁₂	3	5	4	1
P ₁ , P ₂ , F ₁ , F ₂ , RC ₁₁ , RC ₁₂	3	6	4	1

3.4. Testes de Hipóteses

De posse das funções de verossimilhança para cada modelo, é possível então compor testes de interesse, por razão de verossimilhanças, considerando diferentes hipóteses. Por exemplo, confrontando os modelos 1 e 5 (Tabela 5), testa-se a hipótese de herança monogênica e poligênica contra a hipótese de herança poligênica somente.

Tais testes de razão de verossimilhanças são feitos por meio da estatística LR (Mood et al., 1974).

De maneira geral a estatística LR é dada por:

$$LR = -2 \ln \frac{L(M_i)}{L(M_j)} \quad (1.24),$$

sendo $L(M_i)$ e $L(M_j)$ as funções de verossimilhança dos modelos i e j , em que o modelo i deve estar hierarquizado ao modelo j , ou seja, ser um caso particular seu. Esta estatística segue uma distribuição aproximadamente de qui-quadrado em que, para um teste com aproximação α , quando H_0 é verdadeira, é determinado por:

$$\text{Rejeita-se } H_0 \Leftrightarrow -2 \ln \frac{L(M_i)}{L(M_j)} > \chi^2_{(1-\alpha, v)},$$

em que v é o número de graus de liberdade dado pela diferença entre os números de parâmetros dos modelos M_j e M_i . Para o exemplo citado acima, isto é, confrontado os modelos 1 e 5, a LR é dada por:

$$LR = -2 \ln \frac{L(M_5)}{L(M_1)},$$

pois o modelo 5 é um caso particular do modelo 1 e cujo grau de liberdade é 2. A hierarquização dos modelos ocorre conforme o diagrama da Figura 1, em que M_1, M_2, \dots, M_9 são, respectivamente, o modelo 1, modelo 2, ... modelo 9, mostrados na Tabela 5. Na Figura 1 tem-se, por exemplo, que o modelo M_9 está hierarquizado ao modelo M_8 que estão hierarquizados ao modelo M_3 , que, por sua vez, estão hierarquizados ao modelo mais completo M_1 .

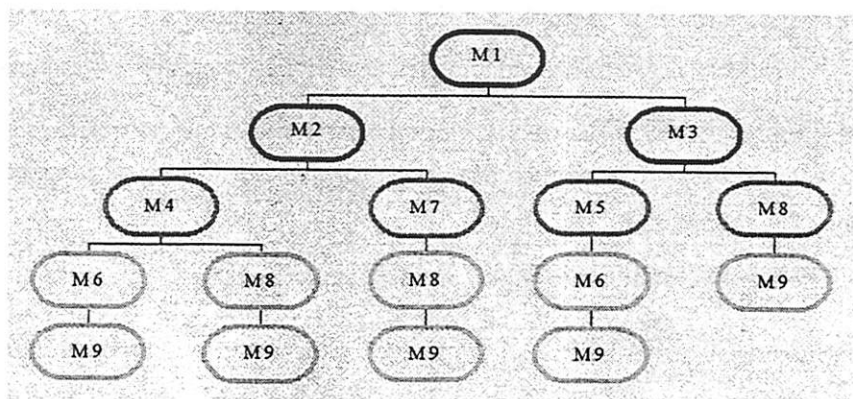


FIGURA 1 - Hierarquização e modelos genéticos.

Por meio destes modelos hierarquizados, outros testes de interesse podem ser estabelecidos, tais como o que confronta os modelos 1 e 7, o que se trata do teste de hipótese de herança monogênica e poligênica versus a hipótese de herança monogênica somente, bem como o que confronta os modelos 7 e 8, em que se testa a hipótese de herança monogênica aditiva com dominância contra a hipótese de herança monogênica aditiva somente.

3.5 Software Estatístico para Estudos de Herança

Como as equações de máxima verossimilhança não resultaram expressões explícitas, foi desenvolvido um software estatístico para estudos de herança, denominado “Monogen”, que implementa as metodologias de estimação de máxima verossimilhança e testes de hipóteses. Este programa utiliza um método numérico iterativo que combina os métodos Quase-Newton e de Powell, um programa de manuseio simples, em que o usuário faz a opção pelo modelo genético de interesse e, automaticamente, os modelos hierarquizados a este são selecionados. Por exemplo, ao selecionar o modelo 5, os modelos hierarquizados 6 e 9, mostrados na Figura 1, são selecionados e automaticamente analisados. Na Figura 2 são mostrados os modelos na ordem decrescente de hierarquização.

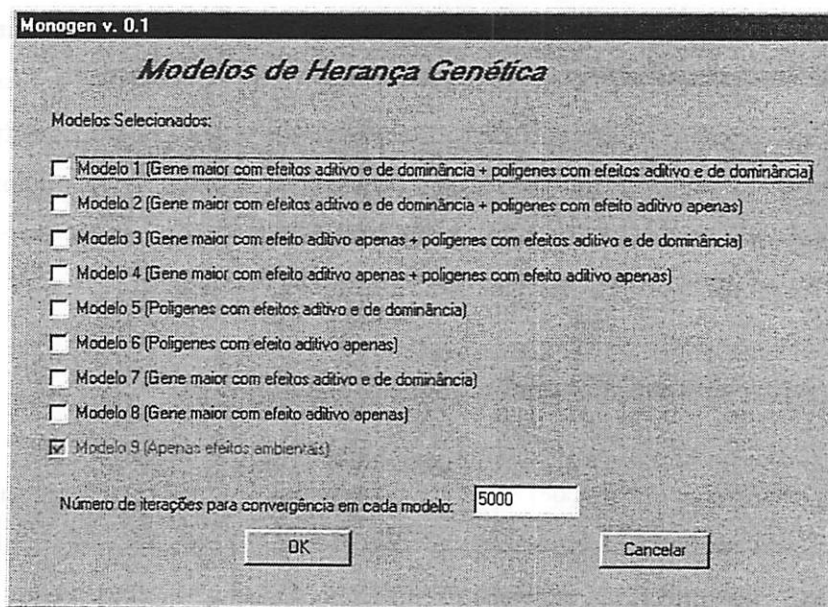


FIGURA 2 – Interface do programa ‘Monogen v. 0.1’ apresentando os modelos de herança genética em ordem hierárquica decrescente.

Seleccionados os modelos de interesse, o programa automaticamente efetua as análises dos dados, bem como os possíveis testes de hipóteses por razão de verossimilhança.

A partir das derivadas segundas, o programa constrói intervalos de confiança, com base na aproximação normal dos estimadores de máxima verossimilhança.

4 EXEMPLO DE APLICAÇÃO

Para ilustrar a metodologia, utilizou-se um conjunto de dados oriundo de um estudo de herança do fenômeno da partenocarpia em abobrinha (*Cucurbita pepo*). Um experimento foi realizado para tal, no setor de olericultura da Universidade Federal de Lavras-MG, considerando as seis populações descritas anteriormente, isto é, P_1 , P_2 , F_1 , RC_{11} , RC_{12} e F_2 em que os genitores contrastantes considerados foram a variedade *Caserta* (P_1) e a variedade *Whitaker* (P_2). O número de plantas amostradas pode ser visto na Tabela 7.

TABELA 7 - Número de plantas amostras por geração

Geração	Número de Plantas Amostradas
P_1	94
P_2	51
F_1	53
RC_{11}	86
RC_{12}	75
F_2	203

A avaliação caracterizou-se por notas de 1 a 5, conforme a maior ou menor ocorrência de partenocarpia, tomando-se médias (ver Anexo, Tabela 1A) das notas por planta atribuídas por três avaliadores.

4.1 Testes de Hipóteses de Modelos Genéticos Hierarquizados de Interesse

Os possíveis testes de hipótese dos modelos genéticos hierarquizados, feitos por meio do programa Monogen v. 0.1, estão apresentados na Tabela 8.

TABELA 8 – Testes de hipóteses de modelos genéticos hierárquicos

Testes	χ^2_c	Graus de liberdade	Prob.
Modelo 1 vs. Modelo 2	9,311445	3	0,0254242
Modelo 1 vs. Modelo 3	0,135244	1	0,7130552
Modelo 1 vs. Modelo 4	11,766019	4	0,0191788
Modelo 1 vs. Modelo 5	10,024717	5	0,0066552
Modelo 1 vs. Modelo 6	20,541712	6	0,0009887
Modelo 1 vs. Modelo 7	1,396269	5	0,9247210
Modelo 1 vs. Modelo 8	12,071388	6	0,0603951
Modelo 1 vs. Modelo 9	230,026822	7	0,0000008
Modelo 2 vs. Modelo 4	2,455647	1	0,1171835
Modelo 2 vs. Modelo 6	11,230258	2	0,0036424
Modelo 2 vs. Modelo 7	Valor negativo*	2	-
Modelo 2 vs. Modelo 8	2,759934	3	0,4301380
Modelo 2 vs. Modelo 9	220,715368	4	0,0000007
Modelo 3 vs. Modelo 5	9,889473	1	0,0016622
Modelo 3 vs. Modelo 6	20,406468	4	0,0004152
Modelo 3 vs. Modelo 8	11,936144	5	0,0356733
Modelo 3 vs. Modelo 9	229,891578	6	0,0000009
Modelo 4 vs. Modelo 6	8,775693	1	0,0030526
Modelo 4 vs. Modelo 8	0,305369	2	0,8584003
Modelo 4 vs. Modelo 9	218,260803	3	0,0000007
Modelo 5 vs. Modelo 6	10,516995	3	0,0146461
Modelo 5 vs. Modelo 9	220,002105	5	0,0000007
Modelo 6 vs. Modelo 9	209,485110	2	0,0000007
Modelo 7 vs. Modelo 8	10,675119	1	0,0010858
Modelo 7 vs. Modelo 9	228,630553	2	0,0000006
Modelo 8 vs. Modelo 9	9,311454	1	0,0022771

* Valor negativo, provavelmente devido a problema de convergência.

Dentre os testes acima, é interessante destacar:

a) Entre os modelos 1 e 5 e entre os modelos 1 e 7, que confrontam a existência de gene de efeito maior mais poligenes com apenas poligenes, ou com apenas gene de efeito maior, respectivamente. Pelo teste da razão de

verossimilhança generalizada para os modelos 1 e 5 rejeita-se H_0 , logo há evidência de que existe um gene de efeito maior.

Pelo teste de hipóteses para os modelos 1 e 7, aceita-se H_0 , isto é, não há evidência de efeitos poligênicos.

b) Assim, é razoável testar os modelos 7 e 8. Pelo teste da razão de verossimilhança generalizada, rejeita-se H_0 . Conclui-se, portanto, que a partenocarpia tem herança monogênica dominante e o modelo que melhor se ajusta é o modelo 7, por ser o mais simples e que traduz as informações dos dados. Na Tabela 9 são apresentadas as estimativas dos modelos 1, 5 e 7.

TABELA 9 – Número de iterações, estimativas dos parâmetros μ , A, D, [a], [d], V_A , V_D , S_{AD} , σ^2 e seus respectivos intervalos de confiança (I.C.), relativos aos modelos genéticos hierarquizados 1, 5 e 7

Variável	Modelo		
	1	5	7
Iterações	89	153	85
μ	2,793309	2,797977	2,854684
I.C.(95%)	[2,671475; 2,915142]	[2,650297; 2,945666]	[2,747336; 2,962032]
A	1,297340	-	1,276216
I.C. (95%)	[1,244583; 1,350097]	-	[1,154402; 1,398030]
D	0,209154	-	0,362283
I.C. (95%)	*	-	[0,198067; 0,526500]
[a]	-0,058080	1,260387	-
I.C. (95%)	[-0,216464; 0,100303]	[1,116686; 1,404088]	-

“... continua...”

“TABELA 9, Cont.”

Variável	Modelo		
	1	5	7
[d]	0,267335	0,419250	-
I.C. (95%)	*	[0,133640; 0,704860]	-
V _A	0,000000	0,454944	-
I.C.(95%)	[-0,071166, 0,071166]	[-0,021374; 0,933263]	-
V _D	0,000000	-	-
I.C.(95%)	[-0,086485; 0,086485]	-	-
S _{AD}	0,026791	-	-
σ^2	0,798249	0,839514	0,805040
I.C. (95%)	[0,711930; 0,884569]	[0,648662; 1,030365]	[0,592195; 1,017886]

* Intervalos não construídos devido à ocorrência de estimativas negativas de variâncias do estimador em questão.

5 CONCLUSÕES

Em atendimento aos objetivos propostos neste trabalho tem-se que:

a) os estimadores de máxima verossimilhança dos componentes de média e de variância obtidos são expressões recursivas que requereram o uso de métodos numéricos iterativos;

b) o critério da razão de verossimilhança foi utilizado satisfatoriamente, como o critério para testar as hipóteses de interesse;

c) para a obtenção das estimativas e do teste de hipóteses pelo critério da razão de máxima verossimilhança, um software de análise foi desenvolvido, o Monogen versão 0.1, integrando, em sua composição, o uso da combinação dos métodos numéricos iterativos Quase-Newton e o de Powell;

d) a metodologia foi ilustrada com um conjunto de dados, no qual, pelos testes propostos, verificou-se que o modelo que melhor se ajusta aos dados é o modelo 7 e que a partenocarpia em abobrinha, considerando as variedades estudadas *Caserta* e *Whitaker*, é devida à herança monogênica com ação de dominância.

REFERÊNCIAS BIBLIOGRÁFICAS

- AITKIN, M.; WILSON, G. T. Mixture Models, Outliers, and the EM Algorithm. **Technometrics**, Alexandria, v. 22, n. 3, p. 325-331, Aug. 1980.
- ARIAS, C. A. A.; TOLEDO, J. F. F.; YORINORI, J. T. An improved procedure for testing theoretical segregation models in qualitative genetic studies of soybeans. **Revista Brasileira de Genética**, Ribeirão Preto, v. 17, n. 3, p. 291-297, set. 1994.
- ARIAS, C. A. A.; YORINORI, J. T.; TOLEDO, J. F. F.; KIIHL, R. A. S. Inheritance of resistance of soybean [*Glycine max* (L.) Merrill] to races 4 and 15 of frogeye leaf spot fungus (*Cercospora sojina* Hara). **Brazilian Journal of Genetics**, Ribeirão Preto, v. 19, n. 2, p. 295-304, June 1996.
- AZEVEDO, S. M. Herança da resistência ao vírus da mancha anelar do mamoeiro-estirpe melancia (PRVS-W) em melancia [*Citrullus lanatus* (Trunb.) Matsumara & Nakai]. 2001. 53 p. Tese (Doutorado em Agronomia) - Universidade Federal de Lavras, Lavras, MG.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm (with Discussion). **Journal of the Royal Statistics Society, Serie B**, London, v. 39, n. 1, p. 1-38, 1977.
- FREITAS, J. A.; MALUF, W. R.; CARDOSO, M. G.; GOMES, L. A. A.; BEARZOTI, E. Inheritance of foliar zingiberene contents and their relationship to trichome densities and whitefly resistance in tomatoes. **Euphytica**, Wageningen, v. 127, n. 2, p. 275-287, 2002.

GOMES, L. A. A. **Herança da resistência da alface (*Lactuca sativa L.*) cv. Grand Rapids ao Nematóide de Galhas *Meloidogyne incógnita* (Kofoid & White) Chitwood.** 1999. 70 p. Tese (Doutorado em Agronomia) - Universidade Federal de Lavras, Lavras, MG.

KENNEDY, W. J. Jr.; GENTLE, J. E. **Statistical computing.** New York, 1980. v. 33, 591 p.

LINCH, M.; WALSH, B. **Genetics and analysis of quantitative traits.** Sunderland: Sinauer Associates, 1998. 980 p.

MATHER, K.; JINKS, J. L. **Introdução à genética biométrica.** Tradutor DUARTE, F. A. M. et. al. Ribeirão Preto – SP: Sociedade Brasileira de Genética, 1984. 242 p.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics.** 3. ed. Tóquio: McGraw-Hill Kogakusha, 1974. 564 p.

RAMALHO, M. A. P.; SANTOS, J. B. dos; ZIMMERMANN, M. J. de O. **Genética quantitativa em plantas autógamas: aplicações ao melhoramento do feijoeiro.** Goiânia: UFG, 1993. 271 p.

RESENDE, J. T. V. **Teores de acilólípicos mediadores da resistência a pragas e sua herança em folíolos de tomateiro, obtidos a partir do cruzamento interespecífico *Lycopersicon esculentum* x *L. pennellii*.** 1999. 56 p. Dissertação (Mestrado em Agronomia) – Universidade Federal de Lavras, Lavras, MG.

SOUZA SOBRINHO, F. **“Herança da reação de resistência à raça 2 de *Meloidogyne incógnita* na pimenta *Capsicum annuum L.* cv. Carolina**

Cayenne". 1998. 57 p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras, MG.

WU, R.; LI, B.; WU, S. S.; CASELLA, G. A Maximum Likelihood-Based Method for Mining Major Genes Affecting a Quantitative Character. **Biometrics**, Raleigh, v. 57, n. 3, p. 764-768. Sep. 2001.

ANEXOS

ANEXO A

Página

TABELA 1A	Dados observados sobre estudo de herança do fenômeno da partenocarpia em abobrinha (<i>cucúrbita pepo</i>), variedades <i>Caserta</i> (P ₁) e <i>Whitaker</i> (P ₂) 56
-----------	--

TABELA 1A - Dados observados sobre estudo de herança do fenômeno da partenocarpia em abobrinha (*Cucurbita pepo*), variedades *Caserta* (P₁) e *Whitaker* (P₂)

Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.
P1	1.333	P1	1.000	P1	1.333	P1	1.667
P1	1.333	P1	1.000	P1	1.333	P2	4.000
P1	1.333	P1	2.000	P1	1.333	P2	2.000
P1	1.667	P1	1.667	P1	1.667	P2	4.667
P1	1.333	P1	1.667	P1	1.000	P2	4.333
P1	1.667	P1	4.000	P1	1.667	P2	5.000
P1	1.000	P1	1.000	P1	1.000	P2	4.000
P1	1.000	P1	1.000	P1	1.000	P2	3.000
P1	1.333	P1	1.000	P1	1.000	P2	4.333
P1	1.333	P1	5.000	P1	1.000	P2	4.333
P1	1.000	P1	2.000	P1	1.333	P2	3.667
P1	2.333	P1	1.000	P1	2.000	P2	3.000
P1	1.000	P1	1.000	P1	1.333	P2	3.000
P1	2.667	P1	1.000	P1	1.000	P2	4.000
P1	1.333	P1	2.667	P1	1.333	P2	1.000
P1	1.000	P1	1.000	P1	1.000	P2	4.000
P1	1.000	P1	1.000	P1	4.333	P2	4.333
P1	1.000	P1	1.000	P1	1.000	P2	4.333
P1	1.000	P1	1.000	P1	1.000	P2	3.667
P1	1.000	P1	1.000	P1	1.000	P2	3.333
P1	2.667	P1	1.333	P1	1.000	P2	4.333
P1	1.667	P1	1.000	P1	1.000	P2	4.000
P1	3.000	P1	1.000	P1	1.000	P2	4.000
P1	2.000	P1	1.000	P1	1.000	P2	5.000
P1	2.000	P1	1.667	P1	1.000	P2	5.000
P1	3.000	P1	1.333	P1	2.667	P2	5.000
P1	1.000	P1	2.333	P1	1.000	P2	4.333
P1	3.333	P1	1.667	P1	1.000	P2	3.333
P1	4.333	P1	1.000	P1	3.667	P2	5.000
P1	1.000	P1	1.333	P1	3.000	P2	4.333
P1	1.333	P1	1.000	P1	1.000	P2	4.667

“... continua ...”

“TABELA 1A , Cont.”

Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.
P2	3.333	F1	1.667	F1	3.000	F2	2.000
P2	4.333	F1	3.667	F1	4.000	F2	1.667
P2	4.000	F1	4.333	F1	5.000	F2	3.000
P2	5.000	F1	2.000	F1	1.000	F2	1.667
P2	5.000	F1	4.333	F1	2.667	F2	1.000
P2	4.333	F1	4.333	F1	3.667	F2	5.000
P2	4.667	F1	4.000	F1	5.000	F2	5.000
P2	4.667	F1	4.667	F1	2.333	F2	3.333
P2	4.667	F1	4.333	F2	4.000	F2	4.333
P2	5.000	F1	1.667	F2	1.667	F2	2.667
P2	4.000	F1	2.333	F2	1.000	F2	3.667
P2	5.000	F1	1.667	F2	1.333	F2	2.667
P2	4.667	F1	4.333	F2	2.000	F2	1.000
P2	4.667	F1	4.000	F2	4.667	F2	4.667
P2	4.667	F1	4.667	F2	3.000	F2	4.667
P2	4.333	F1	3.000	F2	3.000	F2	4.333
P2	4.667	F1	4.000	F2	1.333	F2	3.667
P2	3.667	F1	3.000	F2	4.000	F2	1.333
P2	4.333	F1	3.000	F2	1.667	F2	3.667
P2	4.000	F1	2.333	F2	2.333	F2	3.333
P2	4.333	F1	2.667	F2	1.667	F2	1.667
F1	3.667	F1	3.667	F2	1.333	F2	1.667
F1	4.000	F1	1.000	F2	5.000	F2	2.333
F1	1.000	F1	4.333	F2	2.667	F2	3.333
F1	4.000	F1	2.333	F2	2.333	F2	2.667
F1	4.333	F1	3.333	F2	2.667	F2	3.667
F1	4.000	F1	2.000	F2	4.000	F2	3.000
F1	4.000	F1	4.333	F2	1.000	F2	2.667
F1	5.000	F1	2.333	F2	1.333	F2	1.333
F1	4.667	F1	4.000	F2	1.000	F2	4.333
F1	4.000	F1	2.333	F2	1.000	F2	1.333
F1	3.333	F1	2.667	F2	3.000	F2	1.667
F1	3.667	F1	2.000	F2	1.667	F2	1.000

“... continua ...”

“TABELA 1A, Cont.”

Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.
F2	1.667	F2	2.333	F2	1.000	F2	1.667
F2	4.667	F2	3.000	F2	3.333	F2	3.333
F2	2.667	F2	2.667	F2	1.000	F2	4.333
F2	4.000	F2	2.667	F2	2.667	F2	1.667
F2	3.667	F2	3.333	F2	3.000	F2	2.000
F2	1.667	F2	4.667	F2	3.333	F2	5.000
F2	3.667	F2	5.000	F2	1.333	F2	3.667
F2	1.667	F2	5.000	F2	3.667	F2	1.667
F2	1.667	F2	1.000	F2	1.000	F2	2.667
F2	2.667	F2	2.667	F2	3.667	F2	3.667
F2	3.667	F2	3.000	F2	1.333	F2	3.667
F2	2.000	F2	4.000	F2	4.333	F2	1.667
F2	1.000	F2	2.000	F2	1.000	F2	4.333
F2	1.000	F2	4.333	F2	5.000	F2	4.000
F2	1.333	F2	4.667	F2	1.000	F2	3.667
F2	2.667	F2	5.000	F2	2.000	F2	3.000
F2	2.667	F2	4.000	F2	2.667	F2	5.000
F2	3.333	F2	3.000	F2	4.333	F2	1.667
F2	4.333	F2	2.667	F2	1.333	F2	2.000
F2	4.000	F2	2.000	F2	5.000	F2	3.000
F2	4.667	F2	5.000	F2	4.000	F2	2.000
F2	2.333	F2	4.000	F2	4.667	F2	4.000
F2	4.333	F2	3.667	F2	2.667	F2	4.000
F2	3.000	F2	1.333	F2	4.333	F2	1.000
F2	2.667	F2	3.667	F2	4.000	F2	4.667
F2	2.333	F2	3.333	F2	4.000	F2	2.667
F2	3.333	F2	4.000	F2	1.667	F2	4.667
F2	4.333	F2	4.000	F2	3.333	F2	4.667
F2	3.333	F2	1.667	F2	3.333	F2	1.000
F2	4.000	F2	3.667	F2	2.000	F2	3.000
F2	2.000	F2	1.333	F2	4.000	F2	2.000
F2	2.667	F2	1.000	F2	1.000	F2	4.667
F2	1.667	F2	1.000	F2	4.333	F2	1.000

“... continua ...”

“TABELA 1A, Cont.”

Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.
F2	2.333	RC11	1.667	RC11	2.000	RC12	1.000
F2	1.000	RC11	3.333	RC11	1.000	RC12	4.333
F2	3.333	RC11	1.667	RC11	2.000	RC12	2.000
F2	1.333	RC11	2.333	RC11	1.333	RC12	3.333
F2	2.000	RC11	1.333	RC11	2.333	RC12	2.333
F2	1.000	RC11	1.333	RC11	4.000	RC12	4.333
F2	1.333	RC11	2.000	RC11	3.333	RC12	5.000
F2	5.000	RC11	4.667	RC11	1.000	RC12	5.000
F2	4.000	RC11	2.667	RC11	2.000	RC12	4.333
F2	4.333	RC11	3.000	RC11	1.000	RC12	4.667
F2	5.000	RC11	3.667	RC11	2.000	RC12	3.667
F2	4.333	RC11	1.667	RC11	1.000	RC12	2.667
F2	4.333	RC11	4.333	RC11	1.667	RC12	3.000
RC11	1.000	RC11	4.000	RC11	3.333	RC12	3.000
RC11	1.000	RC11	5.000	RC11	4.333	RC12	3.000
RC11	1.000	RC11	3.333	RC11	4.667	RC12	2.333
RC11	2.333	RC11	1.000	RC11	1.667	RC12	3.333
RC11	1.667	RC11	3.333	RC11	5.000	RC12	2.667
RC11	4.000	RC11	2.333	RC11	2.667	RC12	2.667
RC11	2.333	RC11	1.667	RC11	1.000	RC12	1.667
RC11	2.333	RC11	2.000	RC11	1.000	RC12	3.333
RC11	2.000	RC11	1.667	RC11	4.333	RC12	1.000
RC11	1.000	RC11	2.333	RC11	5.000	RC12	2.667
RC11	1.000	RC11	3.000	RC11	1.000	RC12	3.333
RC11	3.333	RC11	5.000	RC11	4.333	RC12	3.667
RC11	1.667	RC11	2.000	RC11	3.000	RC12	5.000
RC11	2.667	RC11	3.667	RC11	1.000	RC12	4.333
RC11	1.000	RC11	2.667	RC11	2.000	RC12	4.333
RC11	1.000	RC11	4.667	RC11	4.333	RC12	4.667
RC11	2.667	RC11	1.000	RC11	1.667	RC12	2.333
RC11	3.667	RC11	2.000	RC11	1.333	RC12	4.000
RC11	2.667	RC11	4.333	RC11	3.667	RC12	4.667
RC11	2.333	RC11	2.333	RC11	1.000	RC12	2.667

“...continua...”

“TABELA 1A, Cont.”

Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.	Geração	Valor Obs.
RC12	3.667	RC12	2.000	RC12	3.667	RC12	4.000
RC12	3.667	RC12	3.000	RC12	5.000	RC12	5.000
RC12	4.333	RC12	4.667	RC12	4.333	RC12	5.000
RC12	2.667	RC12	1.000	RC12	4.000	RC12	5.000
RC12	3.000	RC12	4.000	RC12	4.667	RC12	4.667
RC12	5.000	RC12	4.667	RC12	3.333	RC12	5.000
RC12	4.000	RC12	2.000	RC12	2.000	RC12	2.333
RC12	3.333	RC12	4.667	RC12	4.333	RC12	3.333
RC12	3.000	RC12	5.000	RC12	5.000	RC12	1.667
RC12	3.000	RC12	3.667	RC12	2.500		
RC12	3.333	RC12	5.000	RC12	2.000		