

**AVALIAÇÃO DA RAZÃO DE
VEROSSIMILHANÇA GENERALIZADA EM
TESTES DE HIPÓTESES SOBRE O CONTROLE
GENÉTICO DE CARACTERÍSTICAS
CONTÍNUAS**

RENATA PIRES GONÇALVES

2004

RENATA PIRES GONÇALVES

**AVALIAÇÃO DA RAZÃO DE VEROSSIMILHANÇA GENERALIZADA
EM TESTES DE HIPÓTESES SOBRE O CONTROLE GENÉTICO DE
CARCTERÍSTICAS CONTÍNUAS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador

Dr. Eduardo Bearzoti

LAVRAS

MINAS GERAIS – BRASIL

2004

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

Gonçalves, Renata Pires

Avaliação da razão de verossimilhança generalizada em testes de hipóteses sobre o controle genético de características contínuas / Renata Pires Gonçalves. -- Lavras : UFLA, 2004.

37 p. : il.

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

**1. Experimentação agrícola. 2. Controle genético. 3. Hipótese.
I. Universidade Federal de Lavras. II. Título.**

CDD-519.5
-630.2195

RENATA PIRES GONÇALVES

**AVALIAÇÃO DA RAZÃO DE VEROSSIMILHANÇA GENERALIZADA
EM TESTES DE HIPÓTESES SOBRE O CONTROLE GENÉTICO DE
CARCTERÍSTICAS CONTÍNUAS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

APROVADA em 05 de fevereiro de 2004

Prof. Marcelo Tavares

UFU

Prof. Ruben Delly Veiga

UFLA

Prof. Daniel Furtado Ferreira

UFLA

Prof. Eduardo Bearzoti

UFLA

(Orientador)

LAVRAS

MINAS GERAIS – BRASIL

Aos meus pais,

João Vitalino Gonçalves e Adriana Rosa Pires Gonçalves,

verdadeiros exemplos de vida.

Dedico

À Di, pela paciência, amor e incentivo; aos meus irmãos Ronaldo, Rejanne, Rogério, Roberta, Rodrigo e Ricardo pelo apoio e amor que a mim dedicam; aos meus sobrinhos, sobrinhas, cunhados, cunhadas e aos demais familiares pelo carinho constante e aos amigos pelo incentivo e apoio em todos os momentos.

Ofereço

AGRADECIMENTOS

A DEUS, por sempre iluminar meus caminhos.

À minha família, início de tudo.

Ao professor Eduardo Bearzoti, pela amizade, atenção, orientação e ensinamentos transmitidos durante o curso.

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 pelo apoio obtido durante a realização deste.

A CNPq, pela concessão da bolsa de estudos que possibilitou a realização deste trabalho.

Ao professor Daniel Furtado Ferreira pelos esclarecimentos e sugestões.

Aos professores do Departamento de Ciências Exatas: Augusto, Delly, Joel, Júlio, Lucas, Luiz Henrique, Marcelo, Mário e Thelma pelo apoio e conhecimentos transmitidos.

Aos professores membros da banca examinadora: Eduardo, Daniel, Delly e Marcelo Tavares pela colaboração e sugestões.

Às funcionárias do Departamento de Ciências Exatas pelos auxílios prestados.

À Di, pelo carinho e dedicação nestes quase dois anos, pois não é fácil enfrentar essas estradas e muito menos assistir aos seminários tão “coerentes com sua área”.

Aos meus amigos e colegas de curso: Eliana, Fabyano, Zé, Elisângela, Janser, Néia, Chico, Leite, Rafa, Vânia, Denis, Carla e Cirillo pela amizade, companheirismo e carinho.

Aos amigos de Viçosa (Dandan, Gusta, Oscar, Bola, Gordinha, Corzinha, Clauclau, Luzinha, Nega, Dani, Xica, Ricardão, Marão, Tel, Lana,

Renatinha, Lúcia, Bichinho) e tantos outros. Todos, mesmo distantes fisicamente; sempre tão presentes espiritualmente.

Aos amigos de BH (Osvaldo, Toco, Nego,...) pelo carinho, incentivo e amor.

Ao Marinho, amigo recente, mas não menos importante, pela amizade tão intensa.

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

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Metodologia proposta por Mather e Jinks (1984)	4
2.2 Metodologia proposta por Changjian et al.1994)	8
3 METODOLOGIA	12
3.1 Comparação Entre Modelos 1 e 5 sob H_0	14
3.1.1 Valores dos parâmetros sob $h_p^2 = 0,2$	16
3.1.2 Valores dos parâmetros sob $h_p^2 = 0,7$	17
3.2 Comparação Entre Modelos 1 e 7 sob H_0	17
3.2.1 Valores dos parâmetros sob $h_M^2 = 0,2$	18
3.2.2 Valores dos parâmetros sob $h_M^2 = 0,7$	18
3.3 Cálculo dos parâmetros para avaliação do poder do teste.....	19
3.3.1 Valores dos parâmetros sob $h_p^2 = 0,2$ e $h_M^2 = 0,2$	21
3.3.2 Valores dos parâmetros sob $h_p^2 = 0,2$ e $h_M^2 = 0,7$	21
3.3.3 Valores dos parâmetros sob $h_p^2 = 0,7$ e $h_M^2 = 0,2$	22
4 RESULTADOS E DISCUSSÃO	23
4.1 Avaliação sob H_0 , diferentes tamanhos de amostra e h^2	23
4.2 Avaliação do poder do teste sob diferentes tamanhos de amostra e h^2	30
5 CONCLUSÕES	34
REFERÊNCIAS BIBLIOGRÁFICAS	36

RESUMO

GONÇALVES, Renata Pires. **Avaliação da razão de verossimilhança generalizada em testes de hipóteses sobre controle genético de características contínuas.** LAVRAS: UFLA, 2004. 37p. (Dissertação – Mestrado em Agronomia/Área de concentração em Estatística e Experimentação Agropecuária)*

Estudos de herança genética em plantas são realizados para caracterizar os efeitos genéticos e verificar a existência de um gene de efeito maior e/ou de genes de pequeno efeito (“poligenes”). Quando a característica de interesse é contínua, a verossimilhança é baseada em modelos de misturas de densidades normais. Uma vez que não há testes exatos evidentes para julgar a existência de um gene de efeito maior, a razão de verossimilhança generalizada é em geral utilizada, considerando a aproximação de qui-quadrado. Este trabalho objetivou avaliar esta estatística de teste através de simulação em computador. Dados foram simulados, considerando particularidades de genealogia típicas de tais estudos e duas condições sob a hipótese de nulidade, ou seja, sem a presença de um gene de efeito maior e sem a presença de genes de pequeno efeito (poligenes), para avaliar o controle do erro tipo I. O poder do teste foi avaliado com ambos presentes. No processo de simulação, foram variados o tamanho de amostra e valores do coeficiente de herdabilidade. Resultados indicaram que, embora a distribuição empírica da razão de verossimilhança se tenha desviado significativamente da distribuição de qui-quadrado, houve controle do erro tipo I, considerando um nível de significância nominal de 5%. O poder é elevado para detectar poligenes e gene de efeito maior, em geral. O poder é baixo para detectar gene de efeito maior quando a proporção da variação genética explicada por este é baixa.

* Comitê de Orientação: Dr. Eduardo Bearzoti – UFLA (Orientador) e Dr. Daniel Furtado Ferreira – UFLA.

ABSTRACT

GONÇALVES, Renata Pires. **Evaluation of the generalized likelihood ratio test of hypotheses on the genetic control of continuous traits.** 2004. 37 p. Dissertation (Master in Statistics and Agricultural Experimentation) – Federal University of Lavras, Lavras, MG *

Studies of genetic control in plants are carried out to characterize genetic effects and detect the existence of a major gene and/or genes of minor effects (polygenes). If the trait of interest is continuous, the likelihood can be constructed based on a model with mixtures of normal densities. Once exact tests are not evident with such models, the likelihood ratio test is generally used, using the chi-square approximation. This work aimed at evaluating such test statistic using computer simulation. Data sets were simulated using generations typical in plant studies, under two conditions of null hypothesis, without a major gene, and without polygenes. The power of the test was evaluated with both types of genes present. Different sample sizes and values of heritability were considered. Results showed that, although the empirical densities of the test statistic departed significantly from a chi-square distribution, under null hypotheses, there was a reasonable control of type I error, with a significance level of 5%. The power of the test was generally high to detect polygenes and major genes. Power is low to detect a major gene only when it explains a low fraction of genetic variation.

*Guidance Committee: Dr. Eduardo Bearzoti – UFLA (Adviser) and Dr. Daniel Furtado Ferreira – UFLA.

1 INTRODUÇÃO

Estudos de herança genética em plantas são realizados para caracterizar os efeitos genéticos associados a características econômicas de interesse. Tais estudos são particularmente úteis em características contínuas, para verificar a existência de um gene de efeito maior e/ou de genes de pequeno efeito (“poligenes”).

Existindo um gene de efeito maior, é comum referir-se à herança como “monogênica”. Em tais casos, é interessante estimar o valor genotípico médio de cada uma das três classes genotípicas, a saber, dos dois homozigotos e do heterozigoto.

A estimação de tais parâmetros, bem como o teste da hipótese de herança monogênica têm, em geral, sido feitos utilizando dados observados em plantas das linhagens contrastantes P_1 e P_2 , da geração F_1 (cruzamento entre P_1 e P_2) e F_2 (cruzamento entre plantas F_1), e dos retrocruzamentos RC_1 (entre indivíduos F_1 e P_1) e RC_2 (entre indivíduos F_1 e P_2). A dificuldade estatística inerente a estes estudos é a existência de misturas de distribuições (em geral normais), para as quais os estimadores não são evidentes.

Um método de estimação adequado para tanto é o da máxima verossimilhança, usado por Lynch e Walsh (1998) e Silva (2003) em problemas de misturas de distribuições em genética.

Silva (2003) descreve um modelo estatístico bastante parametrizado envolvendo misturas de densidades normais, o que dificulta a construção de testes exatos para julgar hipóteses de interesse. Assim, testes de razão de verossimilhanças podem ser empregados para tal, considerando a aproximação de qui-quadrado. Neste caso, algumas hipóteses de nulidade naturalmente de interesse correspondem aos seguintes modelos:

Contendo efeitos ambientais e poligênicos (sem gene de efeito maior).

Contendo efeitos ambientais e um gene de efeito maior (sem genes de pequeno efeito, ou seja, poligenes).

Por se tratar de uma distribuição aproximada, estudos de simulação são interessantes para obter conhecimento acerca de sua qualidade (quanto a níveis de significância reais e poder do teste) sob diferentes circunstâncias, em termos de tamanho da amostra, número de genes e herdabilidade.

Este trabalho teve, assim, por objetivos:

- (i) Avaliar a qualidade do teste da razão de verossimilhanças em termos do controle do erro tipo I (sob H_0 , ou seja, ausência de um gene de efeito maior ou de poligenes).
- (ii) Avaliar o poder do teste, ou seja, a capacidade de detecção de genes de maior efeito e de poligenes, quando estão presentes, sob diferentes condições de tamanho de amostra e magnitude de parâmetros genéticos.

2 REFERENCIAL TEÓRICO

No melhoramento genético, em especial no melhoramento de hortaliças, estudos de herança ou controle genético são relativamente comuns, podendo-se citar a herança de resistência ao nematóide em pimenta (Souza Sobrinho, 1998) e em alface (Gomes, 1999), do teor de açúcares no tomateiro (Resende, 1999) e da resistência ao vírus da mancha anelar em melancia (Azevedo, 2001).

Tais estudos são particularmente úteis em características que, embora contínuas, dão indícios de herança monogênica (gene de efeito maior). Em tais casos, a variabilidade dentro de uma dada classe genotípica é resultado da ação de efeitos de ambiente e/ou de genes menores (poligenes). Assim, é interessante estimar o valor genotípico médio de cada uma das três classes genotípicas, a saber, os dois homozigotos e o heterozigoto.

A estimação de tais parâmetros, bem como o teste da hipótese de herança monogênica, têm, em geral, sido feitos utilizando, como populações, linhagens contrastantes P_1 e P_2 , as gerações F_1 e F_2 , e os retrocruzamentos RC_1 (entre indivíduos F_1 e P_1) e RC_2 (entre indivíduos F_1 e P_2), sendo que os valores genotípicos médios de cada um dos três genótipos são estimados utilizando respectivamente as gerações P_1 , P_2 e F_1 (Souza Sobrinho, 1998; Gomes, 1999; Resende, 1999; Azevedo, 2001).

Entretanto, um procedimento mais eficiente (no sentido de utilizar mais informação contida nos dados) faz uso também das gerações RC_1 , RC_2 e F_2 , uma vez que os três genótipos estão representados nestas populações (dois deles em cada um dos retrocruzamentos e todos os três na F_2). A dificuldade estatística que surge neste caso é o fato de que a essas últimas populações correspondem misturas de distribuições (em geral normais), para as quais os estimadores não são evidentes.

Um método de estimação adequado para tanto é o da máxima verossimilhança, que é usado em problemas de misturas de distribuições em genética (Lynch e Walsh, 1998). Esses autores apresentaram este enfoque para a estimação dos efeitos de um gene maior em uma população em equilíbrio de Hardy-Weinberg, no qual também a frequência desse gene de efeito maior é estimada.

Recentemente, Wu et al. (2001) apresentaram uma generalização deste enfoque, admitindo alelismo múltiplo e ilustraram-no com dados de uma população da espécie arbórea *Populus tremuloides*.

Lynch e Walsh (1998) comentam que é possível construir modelos hierarquicamente, aumentando o número de parâmetros e a complexidade da herança da característica em questão. Assim, podem-se admitir efeitos ambientais somente, acrescentar efeitos de poligenes, e efeitos de um gene de efeito maior. Em posse das funções de verossimilhança para cada modelo, é possível então compor testes de razão de verossimilhança considerando diferentes hipóteses.

Algumas das principais metodologias para o estudo de características contínuas são descritas a seguir.

2.1 Metodologia proposta por Mather e Jinks (1984)

Segundo Mather e Jinks (1984), pela herança dissômica, dois alelos B e b, de um dado loco, podem originar três genótipos BB, Bb e bb e dois parâmetros são necessários para descrever as diferenças na expressão fenotípica destes três genótipos com relação a qualquer caráter que afetem, conforme esquematizado na Figura 1.

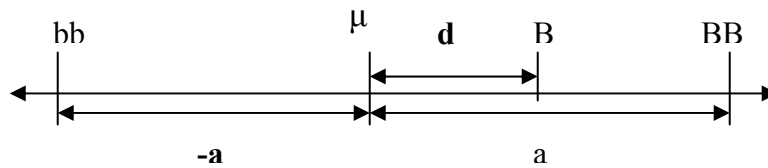


FIGURA 1. Os parâmetros a e d na interação entre os alelos B e b .

Como origem da escala em questão, os autores propõem o valor médio entre os dois homozigotos, representado por μ . Assim a e d são dois parâmetros que medem as diferenças entre os genótipos, sendo que a mede o afastamento de cada homozigoto de μ e d mede o afastamento do heterozigoto da mesma. Tomando B como o alelo que aumenta a expressão do caráter, obtém-se:

- (i) Se $d = 0$, o heterozigoto estará na média e logo não haverá dominância;
- (ii) Se $d > 0$, o heterozigoto estará mais perto de BB , logo B será parcialmente dominante;
- (iii) Se $d = a$, há dominância completa;
- (iv) Se $d < 0$, b será o alelo dominante;
- (v) Se $|d| > a$, o heterozigoto ficará fora dos limites demarcados por BB e bb e o gene será considerado sobredominante.

Quando duas linhagens diferirem em k locos e admitindo que os efeitos destes genes sejam aditivos, define-se $[a]$ como a soma dos efeitos genéticos poligênicos de natureza aditiva, ou seja, a soma de todos os a de cada loco. Da mesma forma, tem-se que $[d]$ é a soma dos efeitos poligênicos de dominância, ou seja, a soma de todos os d .

Considerando o genótipo de um genitor P_1 homozigoto (BB), este terá por expressão $\mu + a$ e o genótipo de um genitor P_2 , também homozigoto (bb), terá por expressão $\mu - a$.

Da mesma maneira, podem-se considerar outros genes. Assim, a média de P_1 é dada por $\mu + [a]$; a média de P_2 corresponde a $\mu - [a]$; o heterozigoto F_1 (Bb), resultado do cruzamento entre indivíduos P_1 e P_2 tem por expressão $\mu + d$ (um gene) e uma média \bar{F}_1 representada por $\mu + [d]$, se forem vários genes. De maneira análoga, constata-se na F_2 , resultante do cruzamento entre indivíduos F_1 , uma média \bar{F}_2 expressa por $\mu + \frac{1}{2}[d]$. Os retrocruzamentos entre indivíduos P_1 e F_1 , P_2 e F_1 , representados por RC_1 e RC_2 respectivamente, têm as expressões de suas médias dadas por $\bar{RC}_1 = \mu + \frac{1}{2}[a] + \frac{1}{2}[d]$ e $\bar{RC}_2 = \mu - \frac{1}{2}[a] + \frac{1}{2}[d]$.

Dessa forma, tem-se os componentes de média com coeficientes dados na Tabela 1.

TABELA 1. Componentes de média, adaptada de Mather e Jinks

	μ	[a]	[d]
P ₁	1	1	0
P ₂	1	-1	0
F ₁	1	0	1
F ₂	1	0	1/2
RC ₁	1	1/2	1/2
RC ₂	1	-1/2	1/2

A pressuposição básica do modelo é a de uma característica contínua com ação de poligenes.

A partir desse modelo básico, pode-se usar o procedimento proposto por Cavalli (1952), denominado teste de escala conjunto. O método consiste em estimar os parâmetros do modelo, μ , [a] e [d] a partir das médias de todos os tipos de famílias disponíveis, seguido por uma comparação dessas médias observadas com seus valores esperados derivados das estimativas dos três parâmetros (Mather e Jinks, 1984).

2.2 Metodologia proposta por Changjian et al. (1994)

As metodologias apresentadas por Changjian et al. (1994) e Silva (2003) apresentam a mesma base teórica e a mesma pressuposição básica, ou seja, característica contínua e ação conjunta de gene de efeito maior e poligenes. Optou-se aqui por descrever a metodologia como apresentada por Silva (2003).

Silva (2003) considerou como modelo geral aquele que apresenta, como características, a existência de um gene de efeito maior mais poligenes, com efeitos aditivos e de dominância e com variância ambiental (σ^2) homogênea.

Admitindo-se distribuição normal, as funções densidade de probabilidade são dadas em (1), por:

$$\begin{aligned}
 P_1 : f_1(y_{i1}) &= \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(y_{i1} - \mu + [a] + A)^2}{2\sigma^2} \right\}, \\
 P_2 : f_2(y_{i2}) &= \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(y_{i2} - \mu - [a] - A)^2}{2\sigma^2} \right\}, \\
 F_1 : f_3(y_{i3}) &= \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(y_{i3} - \mu - [d] - D)^2}{2\sigma^2} \right\}, \\
 RC_1 : f_4(y_{i4}) &= \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D - S_{AD}}} \exp \left\{ -\frac{(y_{i4} - \mu + \frac{[a]}{2} - \frac{[d]}{2} + A)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D - S_{AD})} \right\} \\
 &\quad + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D - S_{AD}}} \exp \left\{ -\frac{(y_{i4} - \mu + \frac{[a]}{2} - \frac{[d]}{2} - D)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D - S_{AD})} \right\},
 \end{aligned} \tag{1}$$

$$\begin{aligned}
RC_2 : f_5(y_{i5}) = & \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D + S_{AD}}} \exp \left\{ -\frac{(y_{i4} - \mu - \frac{[a]}{2} - \frac{[d]}{2} - A)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D + S_{AD})} \right\} \\
& + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 + \frac{V_A}{2} + V_D + S_{AD}}} \exp \left\{ -\frac{(y_{i4} - \mu - \frac{[a]}{2} - \frac{[d]}{2} - D)^2}{2(\sigma^2 + \frac{V_A}{2} + V_D + S_{AD})} \right\},
\end{aligned}$$

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

A partir deste modelo geral (1), nove submodelos podem ser particularizados, conforme a ocorrência ou não de gene de efeito maior com ou sem poligenes e a ocorrência de dominância. Tais modelos estão ilustrados na Tabela 2.

TABELA 2. Modelos genéticos e seus respectivos parâmetros, adaptado de Silva (2003).

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

Tanto em (1) como na Tabela 2, σ^2 designa variância ambiental, μ uma constante de referência; A, o efeito aditivo do gene de efeito maior; D, o efeito de dominância do gene de efeito maior; [a] é o afastamento provocado pelos poligenes da média dos pais homozigotos da linhagem com maior expressão do caráter ou componente poligênico aditivo; [d] é o afastamento devido aos poligenes da média dos pais heterozigotos ou componente poligênico de dominância; V_A é a variância poligênica aditiva; V_D , a variância poligênica de dominância e S_{AD} , o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância (Mather e Jinks, 1984).

Em posse das funções de verossimilhanças para cada modelo, é possível então compor testes de razão de verossimilhanças, considerando diferentes hipóteses.

A função de verossimilhanças é assim definida:

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

sendo que o elemento y_{ij} 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 $P_1, P_2, F_1, RC_1, RC_2$ e F_2).

Aplicando o logaritmo na função L , obtém-se a função suporte (S):

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

Derivando S em relação a cada um dos parâmetros do modelo e igualando as derivadas a zero, obtém-se as equações de máxima verossimilhança.

Os modelos apresentados na Tabela 2 podem ser julgados por testes de razão de verossimilhança apresentados por Silva (2003), que são feitos hierarquicamente, através da estatística LR (Mood et al., 1974):

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

sendo $L(M_i)$ e $L(M_j)$ as funções de verossimilhança dos modelos i e j , sendo que o modelo i deve estar hierarquizado ao modelo j (ou seja, ser um caso particular seu). A estatística LR tem distribuição assintótica de qui-quadrado, com o número de graus de liberdade dado pela diferença entre os números de parâmetros estimados nos modelos i e j .

3 METODOLOGIA

O teste da razão de verossimilhanças estudado neste trabalho é baseado em modelos probabilísticos propostos por Silva (2003) para estudos de herança genética utilizando dados das gerações P₁, P₂, F₁, F₂, RC₁ e RC₂, definidas no item anterior. O modelo mais geral proposto pelo autor admite que cada observação é o resultado de efeitos ambientais, poligênicos e de um gene de efeito maior. Além disso, os efeitos genéticos (tanto maior quanto poligênicos) apresentam dominância, ou seja, um desvio entre o heterozigoto e a média dos homozigotos. A existência de dominância implica em parâmetros adicionais no modelo.

Sob normalidade, há uma única densidade em cada uma das gerações P₁, P₂ e F₁, enquanto que nos retrocruzamentos e na geração F₂ ocorrem misturas de duas e três densidades normais, respectivamente. Representando a densidade normal por $\varphi(\cdot)$, percebem-se as seguintes densidades para cada uma das gerações anteriormente descritas, considerando a existência de poligenes e de um gene de efeito maior:

$$\begin{aligned}
 P_1: & \varphi(\mu_1, \sigma_1^2) & RC_1: & \frac{1}{2}\varphi(\mu_{41}, \sigma_4^2) + \frac{1}{2}\varphi(\mu_{42}, \sigma_4^2) \\
 P_2: & \varphi(\mu_2, \sigma_2^2) & RC_2: & \frac{1}{2}\varphi(\mu_{51}, \sigma_5^2) + \frac{1}{2}\varphi(\mu_{52}, \sigma_5^2) \quad (2) \\
 F_1: & \varphi(\mu_3, \sigma_3^2) \\
 F_2: & \frac{1}{4}\varphi(\mu_{61}, \sigma_6^2) + \frac{1}{2}\varphi(\mu_{62}, \sigma_6^2) + \frac{1}{4}\varphi(\mu_{63}, \sigma_6^2)
 \end{aligned}$$

As médias e variâncias das densidades anteriores, por sua vez, podem ser expressas como:

$$\begin{aligned}
\mu_1 &= \mu - [a] - A & \mu_2 &= \mu + [a] + A \\
\mu_3 &= \mu + [d] + D & \mu_{41} &= \mu - \frac{[a]}{2} + \frac{[d]}{2} - A \\
\mu_{42} &= \mu - \frac{[a]}{2} + \frac{[d]}{2} + D & \mu_{51} &= \mu + \frac{[a]}{2} + \frac{[d]}{2} + A \\
\mu_{52} &= \mu + \frac{[a]}{2} + \frac{[d]}{2} + D & \mu_{61} &= \mu + \frac{[d]}{2} - A \\
\mu_{62} &= \mu + \frac{[d]}{2} + D & \mu_{63} &= \mu + \frac{[d]}{2} + A \\
\sigma_1^2 &= \sigma_2^2 = \sigma_3^2 = \sigma^2 \\
\sigma_4^2 &= \sigma^2 + \frac{V_A}{2} + V_D - S_{AD} \\
\sigma_5^2 &= \sigma^2 + \frac{V_A}{2} + V_D + S_{AD} \\
\sigma_6^2 &= \sigma^2 + V_A + V_D
\end{aligned} \tag{3}$$

Em tais expressões, μ é uma constante de referência; $[a]$ é a soma dos efeitos genéticos poligênicos de natureza aditiva; $[d]$ é a soma dos efeitos poligênicos de dominância; A é o efeito aditivo do gene de efeito maior; D é o efeito de dominância do gene de efeito maior; σ^2 é a variância de natureza ambiental; V_A é a variância aditiva poligênica; V_D é a variância de dominância poligênica; e S_{AD} é o componente da variação relativo ao produto dos efeitos poligênicos aditivos pelos de dominância. A interpretação genética de tais parâmetros é discutida com maior detalhe por Silva (2003) e Mather e Jinks (1984).

Dados foram simulados para as seis gerações, conforme as densidades em (2), e sob H_0 , em duas situações:

- (i) Sem a presença de um gene de efeito maior.
- (ii) Sem a presença de genes de pequeno efeito, ou seja, poligenes.

Para cada uma das situações anteriormente citadas, foi avaliada a qualidade do teste da razão de verossimilhanças em termos do controle do erro tipo I. O poder do teste foi avaliado com ambos presentes (poligenes e gene de efeito maior).

Em uma primeira situação, simularam-se 50, 50, 50, 150, 150 e 200 observações para as gerações P_1 , P_2 , F_1 , RC_1 , RC_2 e F_2 respectivamente e de maneira análoga, simularam-se 25, 25, 25, 75, 75 e 100 observações para as gerações anteriormente citadas, que equivalem a tamanhos de amostras comuns no melhoramento genético. Admitiu-se, sem perda de generalidade, que $\mu = 1000$ e $\sigma^2=1000$. Além disso, estabeleceram-se valores de herdabilidade (a ser definida adiante) iguais a 0,2 e 0,7. A obtenção dos demais parâmetros é descrita a seguir.

3.1 Comparação Entre Modelos 1 e 5 sob H_0

Essa primeira situação corresponde à construção de uma hipótese de nulidade referente à ausência de um gene de maior efeito; para isso atribuiu-se valor igual a zero aos parâmetros A e D.

Optou-se, neste estudo, por admitir dominância completa. Assim, para o poligene i encontra-se $a_i = d_i$.

Seja a variância aditiva V_A expressa por $V_A = \sum_{i=1}^g \frac{a_i^2}{2}$, admitindo-se o número de poligenes $g = 100$ e $a_1 = a_2 = \dots = a_{100} = a$. Assim, constata-se que:

$$V_A = \frac{100}{2}a^2 \text{ e portanto } a = \frac{\sqrt{2V_A}}{10}.$$

A variância de dominância, representada por V_D , é expressa por $V_D = \sum_{i=1}^g \frac{d_i^2}{4}$. Com as pressuposições acima, e substituindo-se d por a , resulta que $V_D = \frac{100}{4}a^2$, logo $V_A = 2V_D$.

Definindo-se h_p^2 como o coeficiente de herdabilidade dos poligenes, sua expressão é dada por $h_p^2 = \frac{V_A + V_D}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$, ou seja, a fração da variância total está sendo explicada pelos poligenes. Substituindo-se $V_A = 2V_D$, $A = D = 0$ e $\sigma^2 = 1000$ na expressão de h_p^2 ocorre:

$$h_p^2 = \frac{2V_D + V_D}{2V_D + V_D + 1000} \Rightarrow 1000h_p^2 + 3V_Dh_p^2 = 3V_D \Rightarrow$$

$$\Rightarrow 3(h_p^2 - 1)V_D = -1000h_p^2$$

logo, as variâncias de dominância e aditiva são dadas respectivamente por

$$V_D = \frac{\sigma^2 h_p^2}{3(1 - h_p^2)} \text{ e } V_A = \frac{2\sigma^2 h_p^2}{3(1 - h_p^2)}.$$

Sendo [a] a soma dos efeitos aditivos dos poligenes, sua expressão é dada por $[a] = \sum_{i=1}^g a_i$; logo $[a] = 100a \Rightarrow [a] = 100 \frac{\sqrt{2V_A}}{10} \Rightarrow$
 $\Rightarrow [a] = 10\sqrt{2V_A}$. Sabendo-se que [d] é a soma dos efeitos de dominância dos poligenes e que $[a] = [d]$ (dominância completa), resulta que:

$$[d] = \sum_{i=1}^g d_i, \text{ logo } [d] = 10\sqrt{2V_A}.$$

Como $[a] = [d]$, constata-se que o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância S_{AD} é dado como sendo o oposto da variância aditiva, ou seja, $S_{AD} = -V_A$.

Considerou-se um valor pequeno e um valor alto para h_p^2 , conforme detalhado a seguir.

3.1.1 Valores dos parâmetros sob $h_p^2 = 0,2$

Para $h_p^2 = 0,2$ verifica-se que as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: $V_A = 166,67$; $V_D = 83,335$; $S_{AD} = -166,67$ e $[a] = [d] = 182,576$.

3.1.2 Valores dos parâmetros sob $h_p^2 = 0,7$

Para $h_p^2 = 0,7$ verifica-se que as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: $V_A = 1555,56$; $V_D = 777,78$; $S_{AD} = -1555,56$ e $[a] = [d] = 557,774$.

3.2 Comparação Entre Modelos 1 e 7 sob H_0

A hipótese de nulidade para essa situação consiste em supor ausência de genes de pequeno efeito (poligenes). Para isso atribuiu-se valor igual a zero para os parâmetros $[a]$, $[d]$, V_A , V_D e S_{AD} .

Definindo-se h_M^2 como o coeficiente de herdabilidade do gene de efeito maior, sua expressão é dada por $h_M^2 = \frac{\frac{A^2}{2} + \frac{D^2}{4}}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$, com variância genética de referência em F_2 , ou seja, $\frac{A^2}{2} + \frac{D^2}{4}$.

Supondo-se que o efeito aditivo e o efeito de dominância do gene de efeito maior são iguais e substituindo $A = D$, $V_A = 2V_D = 0$ e $\sigma^2 = 1000$ obtém-se:

$$h_M^2 = \frac{\frac{A^2}{2} + \frac{A^2}{4}}{\frac{A^2}{2} + \frac{A^2}{4} + \sigma^2} \Rightarrow h_M^2 = \frac{\frac{3}{4}A^2}{\frac{3}{4}A^2 + 1000} \Rightarrow$$

$$\frac{3}{4}A^2 h_M^2 + 1000h_M^2 = \frac{3}{4}A^2 \Rightarrow \frac{3}{4}A^2(1 - h_M^2) = 1000h_M^2$$

$$A^2 = \frac{4}{3} \cdot 1000 \cdot \frac{h_M^2}{1 - h_M^2}.$$

Dessa forma comprova-se que o efeito aditivo do gene de efeito maior é dado por $A = 2\sqrt{\frac{1000h_M^2}{3(1 - h_M^2)}}$ e como $A = D$ tem-se que $D = 2\sqrt{\frac{1000h_M^2}{3(1 - h_M^2)}}$.

Da mesma maneira, dois valores de h_M^2 foram considerados, 0,2 e 0,7.

3.2.1 Valores dos parâmetros sob $h_M^2 = 0,2$

Para $h_M^2 = 0,2$ verifica-se que os efeitos aditivo e de dominância do gene de efeito maior são dados por: $A = D = 18,257$.

3.2.2 Valores dos parâmetros sob $h_M^2 = 0,7$

Para $h_M^2 = 0,7$ verifica-se que os efeitos aditivo e de dominância do gene de efeito maior são dados por: $A = D = 55,777$.

3.3 Cálculo dos parâmetros para avaliação do poder do teste

Neste caso, considerou-se a partição $h^2 = h_M^2 + h_p^2$, sendo h_M^2 o coeficiente de herdabilidade do gene de efeito maior e h_p^2 o coeficiente de herdabilidade dos genes de pequeno efeito (poligenes).

Como já descrito anteriormente, constata-se:

$$h_M^2 = \frac{\frac{A^2}{2} + \frac{D^2}{4}}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$$

$$h_p^2 = \frac{V_A + V_D}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$$

substituindo-se $A = D$ e $V_A = 2V_D$ em h_M^2 e h_p^2 , obtém-se:

$$h_M^2 = \frac{\frac{3}{4}A^2}{\frac{3}{4}A^2 + 3V_D + \sigma^2} \text{ e } h_p^2 = \frac{3V_D}{\frac{3}{4}A^2 + 3V_D + \sigma^2}.$$

Da mesma forma, substituindo-se $V_D = \frac{\sigma^2 h^2}{3(1 - h^2)}$ em h_p^2 origina-se:

$$h_p^2 = \frac{3V_D}{\frac{3}{4}A^2 + 3V_D + \sigma^2} \Rightarrow \frac{3}{4}A^2h_p^2 + 3V_Dh_p^2 + \sigma^2h_p^2 = 3V_D$$

⇒

$$\Rightarrow 3(1-h_p^2)V_D = \frac{3}{4}A^2h_p^2 + \sigma^2h_p^2 \quad \text{e} \quad V_D = \frac{\frac{3}{4}A^2h_p^2 + \sigma^2h_p^2}{3(1-h_p^2)}$$

Substituindo-se $V_D = \frac{\frac{3}{4}A^2h_p^2 + \sigma^2h_p^2}{3(1-h_p^2)}$ em h_M^2 , sabe-se que:

$$h_M^2 = \frac{\frac{3}{4}A^2(1-h_p^2)}{\frac{3}{4}A^2 + 3V_D + \sigma^2} \Rightarrow \frac{3}{4}A^2h_M^2 + \sigma^2h_M^2 = \frac{3}{4}A^2(1-h_p^2)$$

⇒

$$\Rightarrow \frac{3}{4}A^2[h_M^2 - (1-h_p^2)] = -\sigma^2h_M^2 \Rightarrow A^2 = \frac{\sigma^2h_M^2}{\frac{3}{4}[(1-h_p^2) - h_M^2]}$$

Dessa forma verifica-se que o efeito aditivo do gene de efeito maior é dado por:

$$A = 2\sqrt{\frac{\sigma^2h_M^2}{3[1 - (h_p^2 + h_M^2)]}}$$

E substituindo-se $A = 2 \sqrt{\frac{\sigma^2 h_M^2}{3 [1 - (h_p^2 + h_M^2)]}}$ em V_D , obtém-se

que a variância dos efeitos poligênicos de dominância é dado por:

$$V_D = \frac{\frac{\sigma^2 h_M^2 h_p^2}{[1 - (h_p^2 + h_M^2)]} + \sigma^2 h_p^2}{3(1 - h_p^2)} .$$

Consideraram-se combinações de valores 0,2 e 0,7 para h_M^2 e h_p^2 , o que determinou valores específicos para os parâmetros, conforme detalhado a seguir.

3.3.1 Valores dos parâmetros sob $h_p^2 = 0,2$ e $h_M^2 = 0,2$

Para $h_p^2 = 0,2$ e $h_M^2 = 0,2$ sabe-se que os efeitos aditivos e de dominância do gene de efeito maior, as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: $A = D = 21,082$; $V_A = 222,22$; $V_D = 111,11$; $S_{AD} = -222,22$ e $[a] = [d] = 210,817$.

3.3.2 Valores dos parâmetros sob $h_p^2 = 0,2$ e $h_M^2 = 0,7$

Para $h_p^2 = 0,2$ e $h_M^2 = 0,7$ reconhece-se que os efeitos aditivos e de dominância do gene de efeito maior, as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: $A = D = 96,609$; $V_A = 1333,334$; $V_D = 666,667$; $S_{AD} = -1333,334$ e $[a] = [d] = 516,398$.

3.3.3 Valores dos parâmetros sob $h_p^2 = 0,7$ e $h_M^2 = 0,2$

Para $h_p^2 = 0,7$ e $h_M^2 = 0,2$ tem-se conhecimento de que os efeitos aditivos e de dominância do gene de efeito maior, as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: $A = D = 51,639$; $V_A = 4666,666$; $V_D = 2333,333$; $S_{AD} = -4666,666$ e $[a] = [d] = 966,092$.

Especificados valores para os parâmetros, os dados foram simulados de maneira usual, ou seja, sorteando-se valores entre 0 e 1, os quais eram usados na inversa da função distribuição de probabilidade normal. No caso de o dado pertencer a uma geração contendo uma mistura de distribuições (um retrocruzamento ou F_2), então a simulação era feita em duas etapas. Primeiramente, sorteava-se de qual componente da mistura o indivíduo pertencia, conforme probabilidades apropriadas, obedecendo à proporção 1:1 nos retrocruzamentos e 1:2:1 em F_2 . Em seguida, sorteava-se uma realização de variável aleatória normal, correspondente ao componente em questão, pelas técnicas usuais de simulação.

O ajustamento aos modelos foi feito da mesma maneira que Silva (2003), alternando o método numérico de Quase-Newton com o de Powell.

As rotinas computacionais necessárias foram escritas em linguagem Delphi, versão 6.

Cada conjunto de dados propiciou um valor de razão de verossimilhanças, dado por (segundo Mood et al., 1974):

$$LR = -2 \ln \frac{L(M_0)}{L(M_1)}$$

sendo $L(M_j)$ a verossimilhança máxima sob o modelo M_j , com $j = 0$ ou 1 .

4 RESULTADOS E DISCUSSÃO

4.1 Avaliação sob H_0 , diferentes tamanhos de amostra e h^2

Nas tabelas 3 e 4, são apresentados os resultados obtidos para controle do erro tipo I, convencionando-se por “n” o conjunto de tamanhos de amostra iguais a 50, 50, 50, 200, 150 e 150 para as gerações P_1 , P_2 , F_1 , F_2 , RC_1 e RC_2 , respectivamente, e coeficientes de herdabilidade iguais a 0,2 e 0,7. Na Tabela 3 há os resultados com amostra n, enquanto que na tabela 4 estão apresentados os resultados referentes à metade deste tamanho da amostra, ou seja, $n/2$, para os mesmos coeficientes de herdabilidade.

TABELA 3. Resultados sob H_0 , com amostra de tamanho n e diferentes valores de h^2 . Sendo h^2 o coeficiente de herdabilidade.

Modelos sob Comparação	Hipótese de Nulidade	h^2	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de Confiança	Kolmogorov Smirnov
1 versus 5	Só Poligenes	0,2	129	0,00775194	[0,00019624, 0,04243306]	0,7760165**
1 versus 5	Só poligenes	0,7	54	0,03703704	[0,00451736, 0,12747179]	0,5798769**
1 versus 7	Só gene maior	0,2	651	0,00000000	[0,00000000, 0,00459117]	0,3278636**
1 versus 7	Só gene maior	0,7	766	0,02088773	[0,01198498, 0,03369879]	0,2567153**

TABELA 4. Resultados sob H_0 , com amostra de tamanho $n/2$ e diferentes valores de h^2 . Sendo h^2 o coeficiente de herdabilidade.

Modelos sob Comparação	Hipótese de Nulidade	h^2	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de Confiança	Kolmogorov Smirnov
1 versus 5	Só poligenes	0,2	233	0,00429185	[0,00010865, 0,02367915]	0,7745634**
1 versus 5	Só poligenes	0,7	99	0,01010101	[0,00025570, 0,05499635]	0,6261338**
1 versus 7	Só gene maior	0,2	766	0,01436031	[0,00718964, 0,02554845]	0,3415536**
1 versus 7	Só gene maior	0,7	878	0,02505695	[0,01576805, 0,03769155]	0,241019**

Embora se tenham simulado 1000 amostras em cada configuração, nem sempre todas eram aproveitadas em função de o critério de convergência não ser satisfeito, quando eventualmente o limite máximo de 1000 iterações por amostra era atingido. Isto refletiu em diferentes números de amostras válidas (Tabelas 3 e 4). Cada conjunto de valores de LR era submetido ao teste de Kolmogorov-Smirnov, para testar sua aderência à distribuição assintótica esperada de qui-quadrado com 2 graus de liberdade na situação 1 versus 5 e com 5 graus de liberdade na situação 1 versus 7. Além disso, a taxa empírica de rejeição de H_0 foi calculada, obtendo-se um intervalo de confiança exato para esta, baseado na distribuição F, para confrontar com o valor nominal de significância utilizado (5%).

Os resultados da simulação permitiram concluir que os valores de razão de verossimilhanças apresentaram uma taxa de rejeição empírica inferior ao valor nominal do nível de significância (0,05). Pode-se dizer, contudo, que os valores com $h^2 = 0,7$, na comparação entre os modelos 1 e 5, foram estatisticamente iguais ao valor nominal, a julgar pelos intervalos de confiança.

As amostras de tamanho n apresentaram maior taxa de rejeição de H_0 (Tabela 3) em relação às amostras de tamanho $n/2$ (Tabela 4), à exceção da comparação 1 versus 7, sob $h^2 = 0,2$. No entanto, deve-se notar que o número de amostras válidas, neste caso, foi pequeno em relação aos demais da mesma configuração, e assim a taxa real de rejeição de H_0 foi estimada com menor precisão.

De qualquer maneira, a tendência geral de a taxa de rejeição de H_0 se aproximar do nível de significância nominal, com o aumento do tamanho da amostra, foi coerente com o esperado, por se tratar de uma distribuição assintótica de qui-quadrado. É possível que, com amostras ainda maiores, tais taxas se aproximassem ainda mais.

O coeficiente de herdabilidade apresentou uma tendência geral de apresentar uma menor taxa de rejeição de H_0 , quando aquele era igual a 0,2, e uma maior taxa, mais próxima ao valor nominal, com $h^2 = 0,7$ (Tabelas 3 e 4).

Esses resultados em conjunto mostram que condições de menor precisão (menor tamanho de amostra e menor h^2) distanciam as taxas de rejeição de H_0 frente ao valor nominal de significância. No entanto, este distanciamento consistentemente apresentou-se no sentido de reduzir a taxa de rejeição de H_0 . Assim, pode-se dizer que, em condições menos favoráveis quanto à precisão, continua havendo um controle satisfatório do erro tipo I, ainda que à custa de um maior conservadorismo do teste.

Embora as densidades sob H_0 tenham sido estatisticamente diferentes das distribuições esperadas de qui-quadrado, pelo teste Kolmogorov-Smirnov (Tabelas 3 e 4), seria interessante confrontar, ainda que visualmente, as densidades empíricas da estatística LR com essas distribuições esperadas. Tais densidades estão apresentadas nas Figuras 2 a 9. De fato, podem-se observar claras semelhanças com a distribuição de qui-quadrado. Há uma evidente assimetria à direita em todos os casos. Além disso, essa assimetria é menos acentuada na comparação entre os modelos 1 e 7, com 5 graus de liberdade, em relação à comparação entre os modelos 1 e 5 (2 graus de liberdade). Isso foi coerente com o comportamento de uma qui-quadrado, a qual tende à simetria, à medida que o número de graus de liberdade tende ao infinito.

Assim, não obstante a significância do teste de Kolmogorov-Smirnov, as Figuras 2 a 9 apontam para uma consistência do critério de teste utilizado, ainda que aproximada (assintótico).

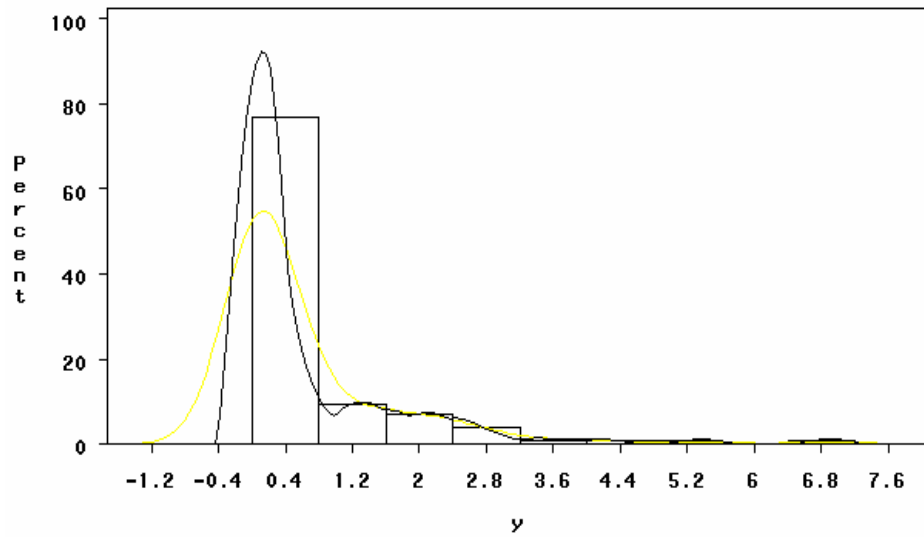


FIGURA 2. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 5, amostra n e $h^2 = 0,2$.

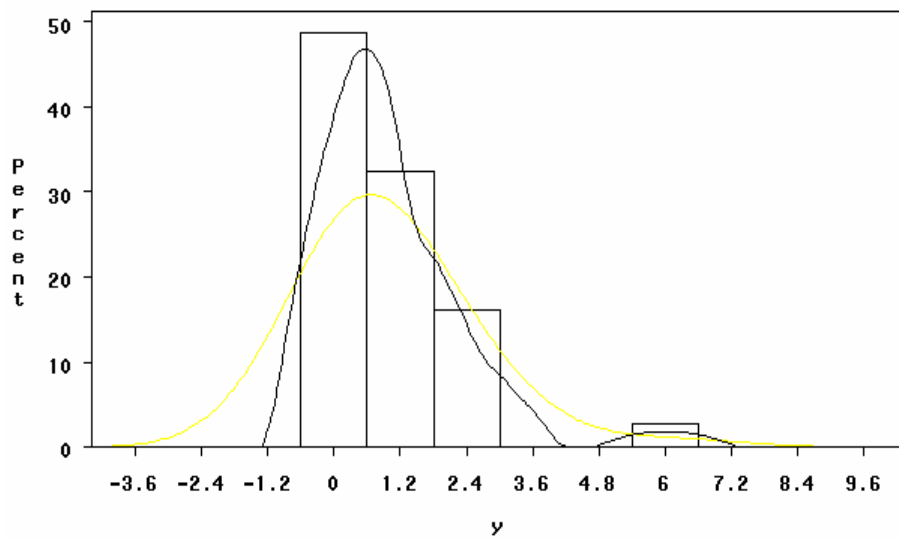


FIGURA 3. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 5, amostra n e $h^2 = 0,7$.

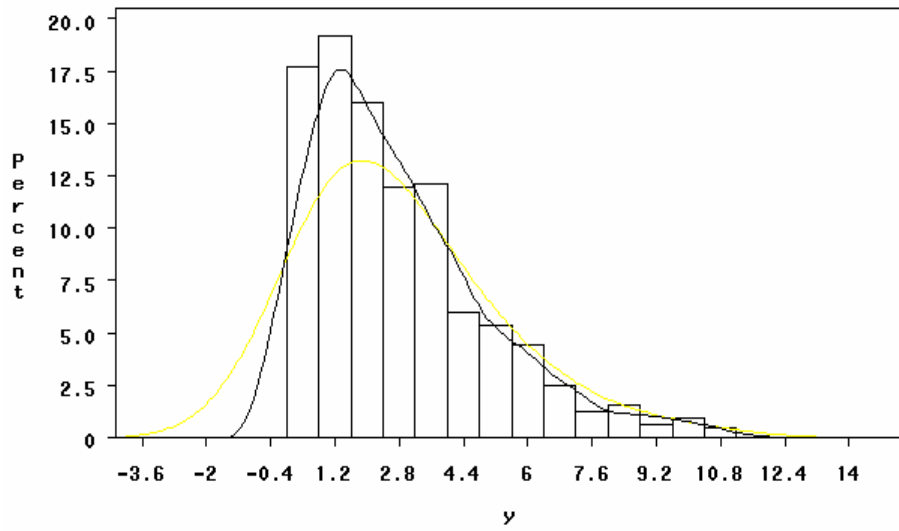


FIGURA 4. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 7, amostra n e $h^2 = 0,2$.

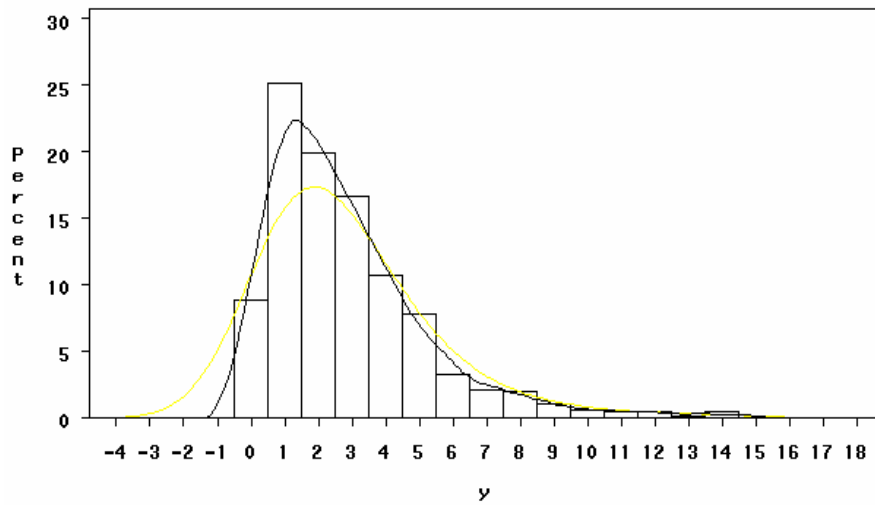


FIGURA 5. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 7, amostra n e $h^2 = 0,7$.

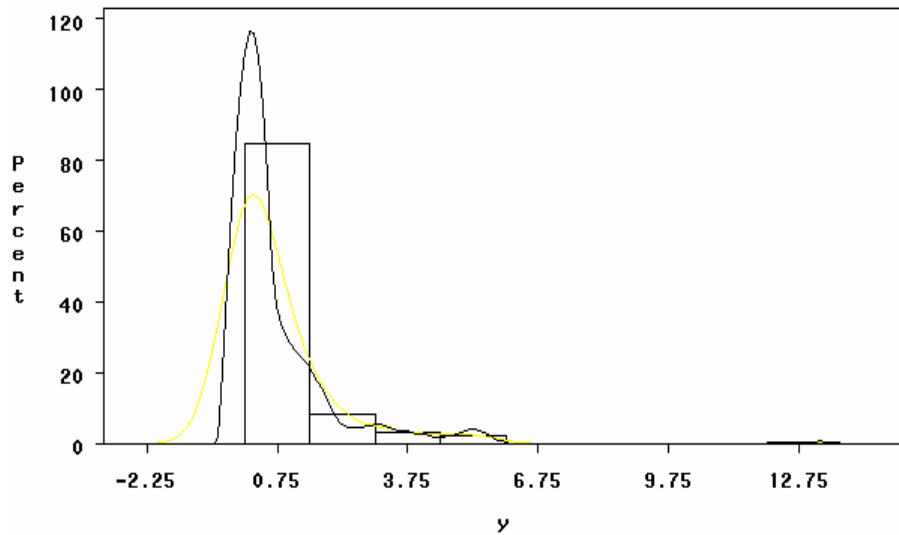


FIGURA 6. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 5, amostra $n/2$ e $h^2 = 0,2$.

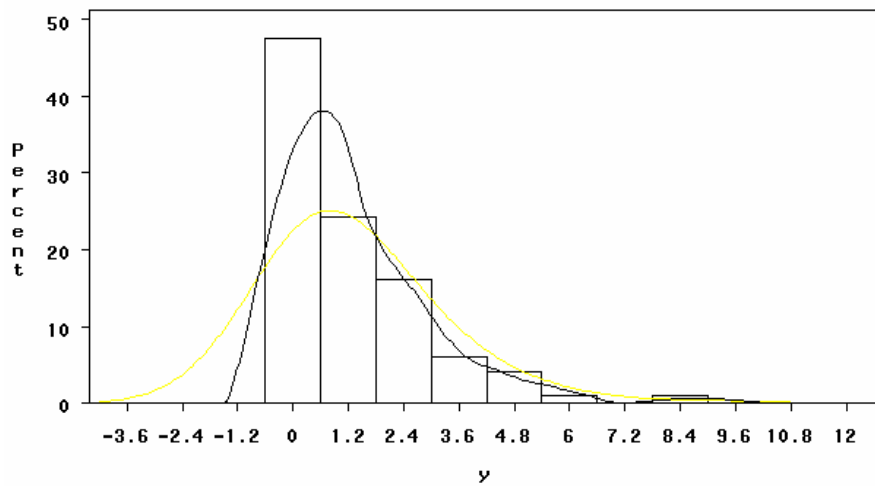


FIGURA 7. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 5, amostra $n/2$ e $h^2 = 0,7$.

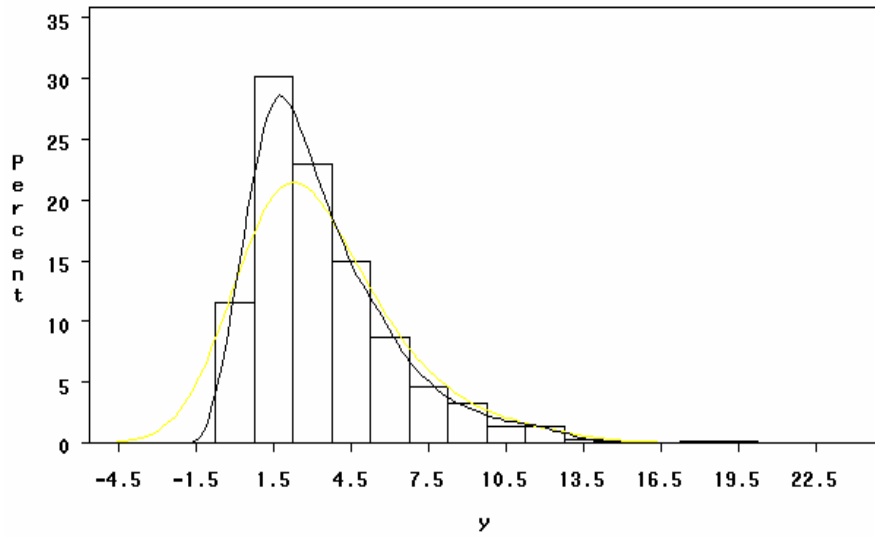


FIGURA 8. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 7, amostra $n/2$ e $h^2 = 0,2$.

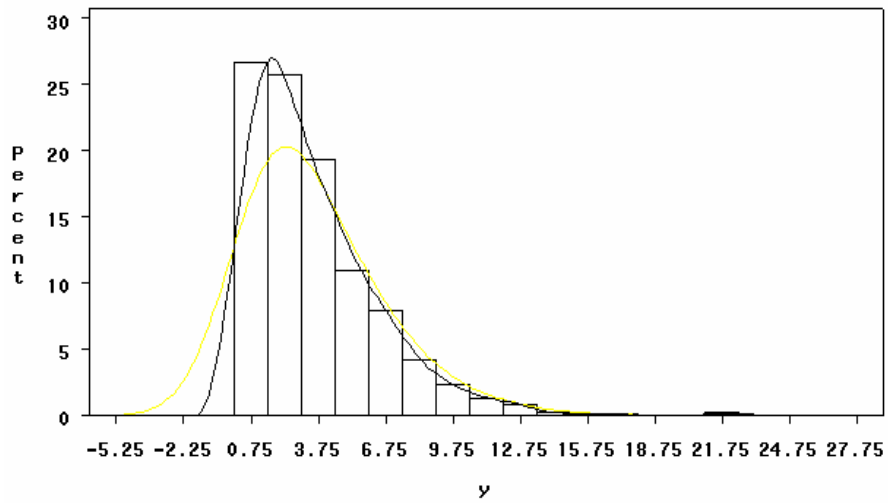


FIGURA 9. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador “Kernel”. Modelo 1 versus 7, amostra $n/2$ e $h^2 = 0,7$.

4.2 Avaliação do poder do teste sob diferentes tamanhos de amostra e h^2

As tabelas 5 e 6 ilustram os resultados obtidos para avaliação do poder do teste. Na Tabela 5, foi utilizado tamanho de amostra n , ou seja, 50, 50, 50, 200, 150 e 150 observações para as gerações P_1 , P_2 , F_1 , F_2 , RC_1 e RC_2 , respectivamente, e uma combinação dos coeficientes de herdabilidade h_M^2 e h_P^2 , iguais a 0,2 e 0,7. Na tabela 6 estão apresentados os resultados referentes à metade do tamanho da amostra acima, ou seja, tamanho $n/2$, para as mesmas combinações entre os coeficientes de herdabilidade.

TABELA 5. Avaliação do poder do teste para amostra de tamanho n .

Modelos sob Comparação	Hipótese de Nulidade	h_P^2	h_M^2	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de Confiança
1 versus 5	Só poligenes	0,2	0,2	626	0,04632588	[0,03124122, 0,06585622]
1 versus 5	Só poligenes	0,2	0,7	415	1,00000000	[0,99280736, 1,00000000]
1 versus 5	Só poligenes	0,7	0,2	160	0,11875000	[0,07302985, 0,17921524]
1 versus 7	Só gene maior	0,2	0,2	761	1,00000000	[0,99607116, 1,00000000]
1 versus 7	Só gene maior	0,2	0,7	518	1,00000000	[0,99423342, 1,00000000]
1 versus 7	Só gene maior	0,7	0,2	174	1,00000000	[0,98293050, 1,00000000]

TABELA 6. Avaliação do poder do teste para amostra de tamanho $n/2$.

Modelos sob comparação	Hipótese de Nulidade	h_p^2	h_M^2	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de Confiança
1 versus 5	Só poligenes	0,2	0,2	614	0,04234528	[0,02784459, 0,06143191]
1 versus 5	Só poligenes	0,2	0,7	409	1,00000000	[0,99270223, 1,00000000]
1 versus 5	Só poligenes	0,7	0,2	159	0,09433962	[0,05376619, 0,15081892]
1 versus 7	Só gene maior	0,2	0,2	840	1,00000000	[0,99644000, 1,00000000]
1 versus 7	Só gene maior	0,2	0,7	632	1,00000000	[0,99527113, 1,00000000]
1 versus 7	Só gene maior	0,7	0,2	199	1,00000000	[0,98505881, 1,00000000]

Percebe-se que também aqui houve uma variação quanto ao número de amostras válidas, em função de nem sempre o critério de convergência ter sido satisfeito, com o número máximo de simulações estipulado (1000).

Nas Tabelas 5 e 6 percebe-se uma tendência geral de alto poder para a estatística de teste. A única exceção verificada foi na detecção do gene de efeito maior, em ambos os tamanhos amostrais, quando a herdabilidade referente àquele era baixa (0,2). A detecção dos poligenes foi significativa em todas as amostras simuladas, em qualquer combinação das herdabilidades h_M^2 e h_p^2 .

O baixo poder para detectar o gene de efeito maior com $h_M^2 = 0,2$ não é realmente um aspecto muito desfavorável, uma vez que, sendo h_M^2 baixa, tal gene tende a assumir um carácter mais de poligene do que propriamente de efeito maior.

Desta maneira, os resultados em conjunto permitem depreender que a estatística de teste, ainda que utilizando um critério de decisão aproximado (assintótico), mostrou-se bastante satisfatória, no sentido que promoveu um controle eficaz do erro tipo I, ainda que conservador, sem com isso acarretar grande prejuízo ao poder estatístico, nas condições avaliadas. Este aspecto torna-se mais expressivo por ter sido verificado em amostras de tamanho inferior ($n/2$) aos geralmente praticados pelos melhoristas de plantas.

A qualidade observada do critério de decisão assintótico é ainda mais expressiva quando se leva em conta o tempo computacional relativamente longo para se obter convergência, o que dificulta a implementação de métodos potencialmente mais adequados, como métodos de reamostragem (bootstrap) ou um enfoque Bayesiano utilizando o amostrador de Gibbs. No entanto, seria interessante estudar a qualidade de tais métodos, futuramente.

Realmente, o aspecto computacional foi um dos fatores mais limitantes neste estudo, uma vez que cada configuração demandava um tempo computacional entre 50 e 168 horas, em geral. Isso limitou o número de configurações estudadas. Seria interessante, contudo, em estudos futuros, avaliar valores mais abrangentes de herdabilidade e de tamanhos de amostra. Outro fator que poderia ser estudado é o controle do erro tipo I e o poder de teste para identificar a existência de interações alélicas de dominância, seja nos poligenes, seja no gene de efeito maior.

É importante ressaltar que o baixo número de amostras válidas observado em diversas situações não significa que existiria uma alta probabilidade de não-convergência em uma situação real, conforme pode parecer à primeira vista. Esse baixo número foi consequência de o número máximo de iterações estipulado ser limitado (1000). Caso contrário o tempo computacional seria muito elevado, por serem muitas as mostras simuladas. Em uma situação real, por se tratar de um único conjunto de dados, é perfeitamente

viável estipular um número máximo bem maior de iterações, aumentando sobremaneira a probabilidade de convergência. Isso foi feito em algumas condições e tal comportamento pôde ser verificado (dados não apresentados).

5 CONCLUSÕES

- (i) Os resultados da simulação mostraram que a densidade de razão de verossimilhanças, sob H_0 , desvia-se significativamente da distribuição assintótica de qui-quadrado.
- (ii) Apesar da conclusão (i), pôde-se verificar que o teste da razão de verossimilhanças propiciou um controle adequado do erro tipo I, sendo considerado rigoroso.
- (iii) No controle do erro tipo I, em todos os casos, a taxa de rejeição empírica de H_0 esteve abaixo de 0,05.
- (iv) Em geral coeficientes de herdabilidade mais baixos ($h^2 = 0,2$) tenderam a abaixar a taxa de rejeição de H_0 .
- (v) A redução do tamanho da amostra reduziu a taxa de rejeição de H_0 .
- (vi) A detecção de poligenes é feita com poder elevado, em todas as situações estudadas.
- (vii) O gene de efeito maior é detectado com alto poder, a não ser quando a proporção da variação genética por ele explicada é baixa.

- (viii) O teste da razão de verossimilhança utilizando o critério de decisão assintótico é satisfatório para uso em estudos de herança.

REFERÊNCIAS BIBLIOGRÁFICAS

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.

CHANGJIAN, J.; XUEBIAO, P.; MINGHONG, G. The use of mixture to detect effects of major genes on quantitative characters in a plant breeding experiment. **Genetics**, Austin, v. 136, p. 383-394. Jan. 1994.

CAVALLI, L. L. An analysis of linkage in quantitative inheritance. In: REEVE, E. C. R.; WADDINGTON, C. D. (Ed.). **Quantitative Inheritance** London: HMSO, 1952. p. 135-144.

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.

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.

RESENDE, J. T. V. **Teores de açúcares 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.

SILVA, W. P. **Estimador de máxima verossimilhança em mistura de normais: uma aplicação à genética**. 2003. 60 p. Dissertação (Mestrado) 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. L.; LI, B. L.; WU, S. S.; CASELLA, G. A Maximum likelihood-based method for mining major genes affecting a quantitative character. **Biometrics**, Washington, v. 57, n. 3, p. 764-768, Sept. 2001.