

**ERRO TIPO I, PODER DO TESTE E
INTERVALOS DE CONFIANÇA COM
TÉCNICAS DE REAMOSTRAGEM EM
MODELOS DE HERANÇA GENÉTICA
UTILIZANDO MISTURAS DE NORMAIS**

GISELE MARTINS MOURA

2005

GISELE MARTINS MOURA

**ERRO TIPO I, PODER DO TESTE E INTERVALOS DE CONFIANÇA
COM TÉCNICAS DE REAMOSTRAGEM EM MODELOS DE
HERANÇA GENÉTICA UTILIZANDO MISTURAS DE NORMAIS**

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 a obtenção do título de "Mestre".

Orientador

Prof. Dr. Eduardo Bearzoti

**LAVRAS
MINAS GERAIS – BRASIL
2005**

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

Moura, Gisele Martins

Erro tipo I, poder do teste e intervalos de confiança com técnicas de reamostragem em modelos de herança genética utilizando misturas de normais / Gisele Martins Moura. -- Lavras : UFLA, 2005.

69 p. : il.

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

**1. Herança genética. 2. Erro tipo I. 3. Poder de teste. 4. Intervalos de confiança.
5. Técnicas de reamostragem. I. Universidade Federal de Lavras. II. Título.**

CDD-519.52

GISELE MARTINS MOURA

**ERRO TIPO I, PODER DO TESTE E INTERVALOS DE CONFIANÇA
COM TÉCNICAS DE REAMOSTRAGEM EM MODELOS DE
HERANÇA GENÉTICA UTILIZANDO MISTURAS DE NORMAIS**

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 a obtenção do título de “Mestre”.

APROVADA em 28 de fevereiro de 2005

Prof. Dr. Daniel Furtado Ferreira	UFLA
Prof. Dr. Ruben Delly Veiga	UFLA
Prof. Dr. Heyder Diniz Silva	UFU



Prof. Dr. Eduardo Bearzoti
UFLA
(Orientador)

**LAVRAS
MINAS GERAIS – BRASIL**

Ao Pai, ao Filho e ao Espírito Santo,
Pelo plano de vida,

Louvo e agradeço.

Aos meus pais, Agrimar e Nair, co-responsáveis na construção deste projeto, pelos brilhantes exemplos que são e pelo apoio incondicional, que os fizeram muitas vezes, renunciar a seus sonhos em favor dos meus,

Dedico esta vitória.

Ao meu noivo Lucas, meu grande amor,
pela paciência, dedicação e incentivo.

Ofereço.

AGRADECIMENTOS

A todos meus familiares da “FAMÍLIA VÔ MENDENGO” e não menos importantes, os da “FAMÍLIA VÓ LIA”, pelos maravilhosos exemplos de vida, pelo incentivo, carinho e confiança.

Aos meus irmãos, Humberto e Agnomar, minha sobrinha Monise e minha sogra e amiga, Zulma, pelo carinho e amizade.

Ao Prof. Dr. Eduardo Bearzoti, não somente pela orientação e ensinamentos, mas pela disponibilidade, apoio e amizade.

À Universidade Federal de Lavras, pela oportunidade de realização do curso. Em especial ao Departamento de Ciências Exatas, por todas as condições oferecidas na realização do curso, aos funcionários, pela acolhida simples e carinhosa; aos professores, pela alegre convivência e pelos ensinamentos.

À CAPES, pelo imprescindível apoio financeiro, viabilizando a realização deste trabalho.

Aos meus colegas de curso, pelos bons momentos que passamos juntos, de maneira especial à Imaculada e à Roberta, por todo apoio e carinho.

A todos aqueles que Deus, em sua sabedoria, escolheu com muito carinho para estarem em meu caminho em Lavras, principalmente aos amigos do GPP (Grupo de Partilha e Perseverança - São Benedito), grande “ pilar” nesta etapa de vida. E a tantos outros, que tive o prazer de conviver, o casal Lili e Breno, sempre muito “fofos”; aos padrinhos Leonel e Cláudia, que me fizeram parte da família. E de maneira muito especial, às minhas famílias de república, a princípio, Rafaela e Vânia, e depois na República Trem da Alegria: Samantha, Nádia e Andressa, pela alegre convivência, amizade e carinho.

A todos que acreditaram e colaboraram.... Valeu!!!

BIOGRAFIA

GISELE MARTINS MOURA, filha de Agrimar Martins de Moura e Nair Benedita de Moura nasceu em Tupaciguara, Minas Gerais, em 17 de abril de 1982. Viveu num cantinho de paraíso, na Fazenda Córrego da Gurita (numa linda região conhecida por Descarga) até 1998. No ano seguinte em Uberlândia iniciou sua graduação em Matemática, na Universidade Federal de Uberlândia, UFU.

Durante a graduação, foi bolsista do CNPq, sob a orientação do Professor Ednaldo Carvalho Guimarães, na Faculdade de Matemática. Nesse período, pesquisou a dependência espacial da precipitação pluviométrica no Triângulo Mineiro e Alto Paranaíba.

Concluiu sua licenciatura em 2002, ingressando em seguida no Mestrado em Agronomia (Área de concentração: Estatística e Experimentação Agropecuária) na Universidade Federal de Lavras, UFLA.

SUMÁRIO

	Página
RESUMO	i
ABSTRACT	ii
CAPÍTULO 1	01
INTRODUÇÃO GERAL	02
REFERENCIAL TEÓRICO	04
1 Herança genética	04
1.1 Herança poligênica	07
1.2 Herança monogênica	08
1.3 Modelos de misturas de normais	10
1.4 Grau médio de dominância	15
1.5 Herdabilidade	16
2 Taxa de erros tipo I e II e poder do teste	16
3 Técnicas de reamostragem	17
3.1 Jackknife	17
3.2 Bootstrap	19
3.2.1 Intervalo de confiança percentil simples	20
3.2.2 Intervalo de confiança BCa (correção para o viés e aceleração)	20
REFERÊNCIAS BIBLIOGRÁFICAS	23
CAPÍTULO 2: Avaliação da qualidade do teste de razão de verossimilhança, quanto ao controle do erro tipo I e do poder do teste	26
RESUMO	27
ABSTRACT	28
INTRODUÇÃO	29
MATERIAL E MÉTODOS	30

1 Cálculo dos parâmetros para avaliação do erro tipo I sob H_0	31
1.1 Modelos sob comparação: 1 e 5	31
1.2 Modelos sob comparação: 1 e 7	34
2 Cálculo dos parâmetros para avaliação do poder do teste	36
RESULTADOS E DISCUSSÃO	40
1 Avaliação do erro tipo I sob H_0	40
2 Avaliação do poder	46
CONCLUSÕES	55
REFERÊNCIAS BIBLIOGRÁFICAS	56
CAPÍTULO 3: Técnicas de reamostragem jackknife e bootstrap na estimação por intervalo dos parâmetros genéticos ...	57
RESUMO	58
ABSTRACT	59
INTRODUÇÃO	60
MATERIAL E MÉTODOS	61
RESULTADOS E DISCUSSÃO	63
CONCLUSÕES	67
CONCLUSÕES GERAIS	68
REFERÊNCIAS BIBLIOGRÁFICAS	69

RESUMO

MOURA, Gisele Martins. **Erro tipo I, poder do teste e intervalos de confiança com técnicas de reamostragem em modelos de herança genética utilizando misturas de normais.** Lavras: UFLA, 2005. 68p. (Dissertação - Mestrado em Agronomia/ Estatística e Experimentação Agropecuária).¹

Estudos que têm por objetivo determinar a natureza das interações alélicas e verificar a existência de um gene principal e/ou poligenes, são comuns nos programas de melhoramento genético vegetal. Se a característica de interesse é contínua, as distribuições de probabilidade são misturas de densidades, por exemplo, normais. O teste de razão de verossimilhança pode ser utilizado para verificar a existência de poligenes e/ou gene principal. Para testar a qualidade de tal teste, foram avaliados o erro tipo I e o poder do teste, sob diferentes condições típicas nos estudos de melhoramento genético, utilizando simulação de dados. Os resultados mostraram que o teste de razão de verossimilhança proporcionou um controle adequado do erro tipo I, porém, rigoroso. O poder se mostrou elevado, tanto na detecção de poligenes quanto de gene principal, com exceção da situação de detecção do gene principal, quando a proporção da variação genotípica do mesmo era baixa. Enfim, o teste de razão de verossimilhança utilizando o critério de decisão assintótico mostrou-se uma excelente alternativa nos estudos de herança. Outro objetivo deste trabalho foi obter estimativas intervalares de parâmetros genéticos relativos a poligenes e gene principal. Tais estimativas utilizando técnicas de reamostragem jackknife e bootstrap; no bootstrap, foram calculados o intervalo percentil simples e BCa (correção para o viés e aceleração). Pôde-se concluir que os intervalos são uma boa alternativa para tais estimativas, tendo mostrado uma coerência com as conclusões dos testes da razão de verossimilhança, quanto ao modelo escolhido.

¹ **Comitê Orientador:** Dr. Eduardo Bearzoti – UFLA.(Orientador), Dr. Daniel Furtado Ferreira – UFLA.

ABSTRACT

MOURA, Gisele Martins. Type I Error, Power Test and Confidence Intervals With Resampling Techniques in Genetic Models Using Mixtures of Normal Densities. Lavras: UFLA, 2005. 68p. (Dissertation – Master in Agronomy/ Major in Statistics and Agricultural Experimentation).¹

In plant breeding programs, it is common to carry out studies to investigate the nature of allelic interactions and if a major gene and/or polygenes are present. If the trait under consideration is continuous and a major gene is present, the probability distribution is a mixture of densities, generally Gaussian. The likelihood ratio test can be used to test the existence of polygenes and/or major gene. Data simulation was used to evaluate such test, accounting for the type I error rate and the power of the test. Results showed that the test kept type I error rates under the nominal value, being somewhat conservative. Test power was high to detect polygenes and the major gene in most cases, but not to detect the major gene when it explained a small ration of phenotypic variation. Therefore, the likelihood ratio test was excellent, in general. Another aim of this study was to obtain interval estimates of the parameters using the resampling techniques of bootstrap and jackknife. The bootstrap was implemented using the methods of percentiles and BCa (which has a bias correction and an acceleration parameter). It could be seen that all resampling confidence intervals were adequate, agreeing with the conclusions of the likelihood ratio test with regard the choice of the model.

¹ **Guidance Committee:** Dr. Eduardo Bearzoti – UFLA (Adviser), Dr. Daniel Furtado Ferreira – UFLA

CAPÍTULO 1

ERRO TIPO I, PODER DO TESTE E INTERVALOS DE CONFIANÇA COM TÉCNICAS DE REAMOSTRAGEM EM MODELOS DE HERANÇA GENÉTICA UTILIZANDO MISTURAS DE NORMAIS

INTRODUÇÃO GERAL

Estudos de melhoramento genético em plantas são feitos com frequência para o conhecimento dos efeitos genéticos envolvidos na herança dos caracteres, permitindo determinar e otimizar os métodos de melhoramento.

A maioria das características de interesse econômico apresenta variação contínua, nas quais tem-se interesse em verificar se ocorre um gene principal, com ou sem a ação de poligenes.

Em geral, os estudos de herança têm considerado a estimação dos parâmetros associados aos poligenes (componentes de média e de variância) e ao gene principal separadamente, ou seja, sem a utilização de um modelo único e geral. A inferência dos parâmetros associados aos poligenes é, em geral, feita pela metodologia do teste de escala conjunto; já os parâmetros associados ao gene principal podem ser estimados por meio de testes de aderência.

A estimação dos parâmetros, bem como os testes, tem, em geral, sido feita 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 da geração F_2 (cruzamento entre as plantas de F_1), e dos retrocruzamentos RC_1 (entre indivíduos F_1 e P_1) e RC_2 (entre indivíduos F_1 e P_2).

Changjian et al. (1994) e Silva (2003) descrevem um modelo estatístico bastante parametrizado envolvendo misturas de densidades normais, considerando característica contínua e ação conjunta de gene principal e poligenes. Com este modelo, pode-se avaliar a existência de cada uma destas classes de genes, por meio de testes de razão de verossimilhança, que são feitos com diferentes modelos hierarquizados.

É de interesse avaliar a qualidade de tais testes (quanto a níveis de significância e poder do teste), já que se trata de uma distribuição aproximada, por meio de estudos de simulação, sob diferentes situações, tamanho da amostra

e herdabilidade, conforme feito por Gonçalves (2004). Acredita-se que seria de interesse também avaliar tais testes variando o grau médio de dominância tanto nos poligenes quanto no gene principal, o que não foi considerado por esta autora.

Silva (2003) apresenta ainda intervalos de confiança para tais parâmetros, considerando aproximações normais e erros padrão estimados a partir da matriz de derivadas segundas da função suporte. No entanto, não se conhece a qualidade da aproximação destes erros, nem da distribuição normal dos estimadores. Assim, o ideal seria aplicar técnicas de reamostragem para a obtenção de intervalos de confiança.

No presente trabalho objetivou-se avaliar a qualidade do teste de razão de verossimilhança quanto ao controle do erro tipo I e do poder do teste, sob diferentes condições de tamanho de amostra, herdabilidade e grau médio de dominância, tanto para gene principal quanto para poligenes. Além disso, implementar técnicas de reamostragem na estimação por intervalo dos parâmetros, considerando característica contínua e ação conjunta de gene principal e poligenes.

REFERENCIAL TEÓRICO

1 Herança genética

A estimação de parâmetros em modelos de herança possibilita a obtenção de informações referentes à natureza da ação gênica envolvida na herança dos caracteres, permitindo delinear estratégias adequadas no melhoramento de plantas.

As características podem apresentar segregação contínua ou descontínua. Se for descontínua como textura de sementes ou cor de frutos, a característica é denominada qualitativa; neste caso, fazem-se testes de aderência para julgar hipóteses sobre o número de genes e/ou de alelos (testes de alelismo) e a existência de epistasia. Detalhes de tais metodologias podem ser encontrados em Ramalho et al. (2000).

Por exemplo, Rêgo (1997) avaliou a herança da cor do fruto, em tomate, por meio da análise das gerações F_1 e segregantes, obtidas mediante cruzamento entre plantas da cultivar Santa Clara normal e o mutante amarelo. Este autor comprovou com o teste de χ^2 , a hipótese de a herança do fruto ser monogênica, com dominância completa para o gene que confere cor vermelha ao fruto, havendo ausência de herança materna.

No entanto, a maioria das características de interesse econômico, que geneticistas e melhoristas de plantas trabalham, apresenta segregação contínua, como produção de grãos, altura e ciclo vegetativo de plantas, as quais são denominadas quantitativas. Nestas, é de especial interesse verificar se ocorre um gene principal, com ou sem a ação de poligenes.

Se a característica contínua é governada por um gene principal, ela é, por vezes, denominada herança monogênica; neste caso, a variabilidade dentro de cada classe genotípica é resultado de efeitos ambientais. Esta variação, contudo,

pode estar somada ou não à ação de poligenes que, neste caso, recebe o nome de herança poligênica. Havendo apenas poligenes, a herança é dita poligênica. É claro, contudo, que gene principal e poligenes podem coexistir.

Em geral, os estudos de herança têm considerado a estimação dos parâmetros associados aos poligenes (componentes de média e de variância) e ao gene principal separadamente, ou seja, sem a utilização de um modelo único e geral. A inferência sobre os parâmetros associados aos poligenes é, em geral, feita pela metodologia do teste de escala conjunto, cuja descrição pode ser encontrada em Mather e Jinks (1984) e em Ramalho et al. (1993); já os parâmetros associados ao gene principal podem ser estimados, por exemplo, segundo a metodologia sugerida inicialmente por Souza Sobrinho (1998). Alguns trabalhos com este enfoque podem ser citados.

Souza Sobrinho (1998) fez o estudo da herança da resistência da pimenta ‘Carolina Cayenne’ à raça 2 de *Meloidogyne incognita*, verificando que a herança é monogênica, com ação gênica predominantemente aditiva. Para tal estudo, o autor utilizou as gerações F_1 , F_2 , RC_1 e RC_2 , obtidas a partir dos cruzamentos entre as cultivares ‘Carolina Cayenne’ (P_1) e ‘Agrônômico-8’ (P_2).

Lima Neto (2001) estudou a herança da resistência em tomateiro à murcha bacteriana, a partir das gerações F_1 , F_2 , RC_1 e RC_2 , obtidas com o cruzamento entre as cultivares Drica (P_1) e Santa Clara (P_2). Utilizou, no processo de avaliação de incidência da doença, uma escala de notas de 1 a 5, sendo este um tipo de avaliação que busca representar uma escala contínua de resistência. No estudo, a hipótese de herança monogênica foi rejeitada e, baseado nos componentes de média e de variância de poligenes, o autor concluiu que o modelo aditivo-dominante foi adequado para explicar o tipo de ação gênica envolvida, indicando que os efeitos epistáticos não são importantes na expressão da resistência.

Em abóbora, a herança da resistência ao vírus da mancha foi estudada por Oliveira (1999), utilizando um enfoque semelhante e dispondo das mesmas gerações, obtidas a partir do cruzamento entre as cultivares Baiana Tropical (P_1) e Chirimen (P_2). Também aqui a hipótese de herança monogênica foi rejeitada e o modelo aditivo-dominante se mostrou adequado.

O método de máxima verossimilhança pode ser aplicado para fazer a estimação simultânea dos parâmetros referentes a poligenes e ao gene principal, como apresentado por Changjian et al. (1994) e por Silva (2003). Além disso, os autores também apresentam um caminho para testar o modelo mais adequado, cuja aplicação pode ser observada em alguns trabalhos.

Menezes (2003) estudou a herança da resistência ao vírus da mancha anelar em abobrinha, utilizando as gerações P_1 , P_2 , F_1 , F_2 , RC_1 e RC_2 , em que P_1 é a variedade Caserta e o P_2 a variedade Whitaker. Os resultados mostraram que tal herança é controlada por mais de um loco gênico. O modelo aditivo-dominante foi adequado para explicar o controle do caráter.

Num estudo da herança da resistência da melancia ao vírus *Watermelon mosaic*, Bezerra Jr. (2004), avaliando os genitores, a cultivar Crimson Sweet (P_1 -suscetível) e a introdução PI 595201 (P_2 -resistente), bem como as gerações F_1 , F_2 , RC_1 e RC_2 , concluiu que a existência de um gene principal com efeitos aditivos e de dominância, juntamente com poligenes, parece ser o modelo mais adequado para explicar tal herança.

A seguir serão descritas as principais metodologias que consideram a estimação do modo de herança separadamente e também de forma simultânea, além de alguns conceitos importantes para uma melhor contextualização.

1.1 Herança poligênica

Na herança poligênica, a pressuposição básica do modelo é ter característica contínua com ação de poligenes.

Em geral, a inferência é feita estimando-se componentes de variância e componentes de média. Os componentes de variância são estimados utilizando análise de variância, como pode ser visto em Ramalho et al. (1993). Os componentes de média são, em geral, estimados pela metodologia do teste de escala conjunto, que pode ser encontrado em Mather e Jinks (1984) e em Ramalho et al. (1993). Com essas estimativas, são obtidos valores esperados para as médias e, em seguida, esses valores são comparados com os valores observados por um teste de qui-quadrado.

Sejam [a] a soma dos efeitos aditivos dos poligenes, [d] a soma dos efeitos de dominância dos poligenes e μ uma constante de referência. Então, na presença dos poligenes, as expressões para as médias nas gerações ficam:

$$\bar{P}_1 : \mu + [a] \quad ; \quad \bar{P}_2 : \mu - [a] \quad ; \quad \bar{F}_1 : \mu + [d] \quad ; \quad \bar{F}_2 : \mu + \frac{1}{2}[d]$$

$$\overline{RC}_1 : \mu + \frac{1}{2}[a] + \frac{1}{2}[d] \quad ; \quad \overline{RC}_2 : \mu - \frac{1}{2}[a] + \frac{1}{2}[d]$$

A partir desse modelo, o teste de escala conjunto consiste em estimar os parâmetros μ , [a] e [d], por meio do método de mínimos quadrados ponderados, que pode ser encontrado em Draper e Smith (1998).

Tem-se o modelo linear:

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

em que X é a matriz de incidência, θ é o vetor de parâmetros, ε é o vetor de resíduos e Y o vetor de observações. No presente caso, a matriz de variâncias e covariâncias (V) é diagonal (covariâncias nulas), além disso, as gerações têm

variâncias heterogêneas. Assim, V é positiva definida e, conseqüentemente, V^{-1} também o é. Dessa forma, V^{-1} pode ser fatorada:

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

em que L é a matriz cujos elementos da diagonal são as raízes quadradas dos elementos de V^{-1} . Neste caso, L é igual a L' e, assim, multiplicando-se o modelo por L :

$$LY = LX\theta + Le$$

obtém-se um novo modelo homocedástico e de variância igual a I , pois

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

Utiliza-se o sistema de equações normais para estimar θ no novo modelo; admitindo-se ainda que LX tenha posto coluna completo, tem-se:

$$\hat{\theta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

Além disso, os resíduos do novo modelo são estimados por:

$$Le = LY - LX\hat{\theta}, \text{ e a forma quadrática } (Le)'(Le) = e'LL'e = e'V^{-1}e,$$

sob normalidade, tem distribuição de qui-quadrado, com $n-p$ graus de liberdade, sendo n o número de gerações e p o número de parâmetros estimados.

Esta forma quadrática pode ser escrita como:

$$\chi_c^2 = \frac{n_1}{V_{P1}}(\bar{Y}_1 - \hat{Y}_1)^2 + \frac{n_2}{V_{P2}}(\bar{Y}_2 - \hat{Y}_2)^2 + L + \frac{n_6}{V_{RC1}}(\bar{Y}_6 - \hat{Y}_6)^2$$

Que pode ser utilizada para testar a adequação do modelo. Se rejeitada a hipótese de nulidade, isto pode sugerir, por exemplo, a existência de componentes epistáticos de médias.

1.2 Herança monogênica

Em estudos de herança de características contínuas, é comum o pesquisador investigar a existência de um gene principal. Na literatura são

encontradas diferentes metodologias, porém, a maioria delas tem em comum a construção de classes, com as quais testes de qui-quadrado são feitos. Nestes testes confrontam-se as frequências observadas nestas classes com as esperadas, conforme a geração segregante, como apresentado por Arias et al. (1994) e por Souza Sobrinho (1998). Já que elas apresentam a mesma base conceitual, optou-se por apresentar apenas a metodologia proposta por Souza Sobrinho (1998), a qual foi adaptada, dependendo das particularidades de cada estudo, por mais autores, dentre outros Lima Neto (2001), Oliveira (1999) e Freitas et al. (2002). Será apresentada a metodologia conforme descrito por Freitas et al. (2002).

Esses autores estudaram a herança do teor de zingibereno em folhas de tomateiro (característica claramente contínua), utilizando dados das gerações P_1 , P_2 , F_1 e F_2 . Porém, tal metodologia pode ser estendida em estudos que possuem mais gerações, por exemplo, os retrocruzamentos.

A princípio, é determinado um ponto de truncamento (PT) tal que se discriminem os genitores contrastantes. Admite-se que a característica sob estudo tenha distribuição normal e consideram-se diferentes valores possíveis de grau médio de dominância (GMD). São então implementadas as seguintes etapas:

(i) Com as estimativas usuais da média e da variância das gerações P_1 e P_2 , calculam-se as frequências esperadas, segundo a distribuição normal, abaixo e acima do PT, nas gerações P_1 e P_2 .

(ii) Com cada valor específico de GMD (em geral trabalha-se com valores entre -1 e 1, ou ainda intervalos maiores), estima-se a média da geração F_1 por $\bar{F}_1 = (\bar{P}_1 + \bar{P}_2)/2 + \text{GMD}(\bar{P}_1 - \bar{P}_2)$, em que \bar{P}_1 e \bar{P}_2 são as médias das gerações P_1 e P_2 . A variância da geração F_1 é estimada pela variância amostral entre as plantas F_1 . Com esta média e esta variância, são calculadas as frequências abaixo e acima do PT na geração F_1 , a partir da distribuição normal.

(iii) Na geração F_2 , as freqüências esperadas abaixo e acima do 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.

(iv) 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.

(v) Para os diferentes valores de GMD calcula-se, reiteradas vezes, esta estatística de qui-quadrado. Se os valores dessa estatística forem menores que o valor tabelado, então não se deve rejeitar a hipótese de existência de um gene principal. O GMD estimado é aquele que fornece um qui-quadrado mínimo.

Não há um número preciso de graus de liberdade da estatística de qui-quadrado nesta metodologia (detalhes em Silva, 2003), sendo um valor entre 0 e 2. Freitas et al. (2002) utilizaram o valor intermediário 1.

Apesar das particularidades de cada metodologia, elas correspondem essencialmente a testes de aderência. Outro ponto comum é que toda a informação utilizada para o cálculo das freqüências esperadas é obtida das gerações P_1 , P_2 e F_1 . Seria interessante também utilizar as informações contidas em F_2 e nos retrocruzamentos. Para tanto, segundo Silva (2003), o enfoque estatístico a ser adotado seria reconhecer que as distribuições de probabilidade nestas gerações são misturas de densidades, em geral normais.

1.3 Modelo de herança utilizando misturas de normais

No contexto da genética, modelos de misturas de distribuições normais foram propostos para a detecção de um gene principal em uma população F_2 (Loisel et al., 1994). Este trabalho teve por objetivo estudar o comportamento do teste de razão de verossimilhança para fazer tal detecção.

Changjian et al. (1994) também fizeram uso de modelos de misturas de normais, num trabalho que apresenta um modelo com característica contínua com ação de poligenes e de gene principal, simultaneamente, em um estudo de altura da planta em arroz. Silva (2003) utilizou a mesma idéia num estudo de herança do fenômeno de partenocarpia em abobrinha. Com tal modelo, pode-se avaliar a existência de cada uma destas classes de genes, por meio de testes de razão de verossimilhança, feitos com diferentes modelos hierarquizados.

Embora o princípio básico desses dois trabalhos seja o mesmo, o modelo de Silva (2003) considera qualquer grau de dominância do gene principal, enquanto Changjian et al. (1994) admitem dominância completa. Além disso, na parametrização de Silva (2003) estão explicitados componentes da variação genética poligênica, o que não acontece em Changjian et al. (1994). Assim, optou-se por descrever a metodologia como apresentada por Silva (2003).

Silva (2003) considera dados das gerações P_1 , P_2 , F_1 , F_2 , RC_1 e RC_2 , admitindo-se genes independentes (tanto poligenes como gene principal), e que em todas as gerações as variâncias ambientais (σ^2) são iguais.

Denominando-se os alelos do gene principal como B e b, tem-se, segundo Silva (2003), que aos genótipos BB, Bb e bb correspondem a valores genotípicos iguais a $\mu + A$, $\mu + D$ e $\mu - A$, sendo μ uma constante de referência, A o chamado efeito aditivo e D o efeito de dominância.

Com a existência de poligenes, têm-se ainda os componentes adicionais de média [a] e [d], correspondentes à soma dos efeitos poligênicos aditivos e de dominância, respectivamente, e os componentes poligênicos de variância V_A , V_D e S_{AD} , que são a variância aditiva, a variância de dominância e a soma dos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância. Assim, segundo Silva (2003), têm-se, associadas a cada uma das seis gerações, densidades normais com médias e variâncias específicas. As gerações P_1 , P_2 e F_1 tem-se associada uma única distribuição normal e as

gerações RC₁, RC₂ e F₂ correspondem misturas de duas ou três normais, por haver mais de um genótipo presente, quanto ao gene principal.

Considerando-se o modelo de mistura de densidades normais, é possível associar distribuições de probabilidade a cada uma das seis gerações. Se $N(y_{ij}; a, b)$ representa uma densidade normal da variável Y , com média a e variância b , então, a distribuição de probabilidade associada a y_{ij} (observação correspondente ao indivíduo i da geração j) é dada por:

$$P_1: f_1(y_{i1}) = N(y_{i1}; \mu + [a] + A, \sigma^2);$$

$$P_2: f_2(y_{i2}) = N(y_{i2}; \mu - [a] - A, \sigma^2);$$

$$F_1: f_3(y_{i3}) = N(y_{i3}; \mu + [d] + D, \sigma^2);$$

$$F_2: f_4(y_{i4}) = \frac{1}{4}N\left(y_{i4}; \mu + \frac{[d]}{2} + A, \sigma^2 + V_A + V_D\right) + \\ + \frac{1}{2}N\left(y_{i4}; \mu + \frac{[d]}{2} + D, \sigma^2 + V_A + V_D\right) + \\ + \frac{1}{4}N\left(y_{i4}; \mu + \frac{[d]}{2} - A, \sigma^2 + V_A + V_D\right);$$

$$RC_1: f_5(y_{i5}) = \frac{1}{2}N\left(y_{i5}; \mu + \frac{[a]}{2} + \frac{[d]}{2} - A, \sigma^2 + \frac{V_A}{2} + V_D - S_{AD}\right) + \\ + \frac{1}{2}N\left(y_{i5}; \mu + \frac{[a]}{2} + \frac{[d]}{2} + D, \sigma^2 + \frac{V_A}{2} + V_D - S_{AD}\right);$$

$$RC_2: f_6(y_{i6}) = \frac{1}{2}N\left(y_{i6}; \mu - \frac{[a]}{2} + \frac{[d]}{2} + A, \sigma^2 + \frac{V_A}{2} + V_D + S_{AD}\right) + \\ + \frac{1}{2}N\left(y_{i6}; \mu - \frac{[a]}{2} + \frac{[d]}{2} + D, \sigma^2 + \frac{V_A}{2} + V_D + S_{AD}\right);$$

Como pode ser observado, esta metodologia explicita a distribuição de probabilidade referente a cada observação. Assim, é natural que a estimação dos parâmetros seja feita pelo método da máxima verossimilhança, que consiste em

achar valores para os parâmetros que maximizam a probabilidade de a amostra ter acontecido.

Para descrever tal procedimento, define-se a função de verossimilhança

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

Aplicando o logaritmo na função L, tem-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 em seguida igualando essas expressões a zero, obtêm-se as equações de máxima verossimilhança, ou seja, um sistema de 9 equações, que são não-lineares, não sendo possível deixar os parâmetros em evidência (Silva, 2003). Assim, deve-se lançar mão de procedimentos numéricos, iterativos, para a obtenção da estimativa de tais parâmetros. Silva (2003) apresenta um programa de análise estatística que alterna o uso dos métodos quase-Newton e de Powell.

A partir do modelo genético mais geral, contendo os 9 parâmetros, Silva (2003) apresenta alguns submodelos contendo menos parâmetros, de particular interesse, os quais estão listados na Tabela 1.

TABELA 1: Modelos genéticos e seus respectivos parâmetros.

Modelo	Herança (genes envolvidos)	Gene principal	Poligenes	PARÂMETROS
1	Principal+poligenes	adit. e dom.	adit. e dom.	$\mu, A, D, [a], [d], V_A, V_D, S_{AD}, \sigma^2$
2	Principal+poligenes	adit. e dom.	adit.	$\mu, A, D, [a], V_A, \sigma^2$
3	Principal+poligenes	adit.	adit. e dom.	$\mu, A, [a], [d], V_A, V_D, S_{AD}, \sigma^2$
4	Principal+poligenes	adit.	adit.	$\mu, A, [a], V_A, \sigma^2$
5	Só poligenes	—	adit. e dom.	$\mu, [a], [d], V_A, V_D, S_{AD}, \sigma^2$
6	Só poligenes	—	adit.	μ, A, V_A, σ^2
7	Só gene principal	adit. e dom.	—	μ, A, D, σ^2
8	Só gene principal	adit.	—	μ, A, σ^2
9	Nenhum	—	—	μ, σ^2

Tais modelos podem ser julgados por teste da razão de verossimilhança, por meio da estatística LR (Mood et al., 1974) dada por: $LR = -2 \ln \frac{L(M_i)}{L(M_j)}$, sendo $L(M_i)$ e $L(M_j)$ as funções de verossimilhança dos modelos i e j , em que o modelo i deve ser um caso particular do modelo j .

Esta estatística tem uma distribuição aproximada de qui-quadrado. Um teste com aproximação α , quando H_0 é verdadeira, é determinado pela regra de decisão de rejeitar H_0 se, e somente se, $LR > \chi^2_{(1-\alpha, v)}$. O número de graus de liberdade v é dado pela diferença entre os números de parâmetros dos modelos i e j . No entanto, a rigor, o número de graus de liberdade para o teste utilizando a estatística pode não ser um valor inteiro, quando as proporções de cada termo da mistura são conhecidas (Loisel et al., 1994).

Por exemplo, confrontando os modelos 1 e 5, está-se testando a hipótese de existência de um gene principal. Se a estatística LR for superior ao valor tabelado de qui-quadrado com 2 graus de liberdade, então rejeita-se o modelo 5 e conclui-se que há um gene principal. Também é de especial interesse o teste envolvendo os modelos 1 e 7, com 5 graus de liberdade, que refere-se à hipótese de existência de poligenes. Outros testes com submodelos menores podem também ser definidos.

Por se tratar de uma distribuição aproximada, é de interesse avaliar ainda a qualidade de tais testes (quanto a níveis de significância e poder do teste). Isso foi feito por Gonçalves (2004), fazendo simulação de dados para as seis gerações, conforme as densidades acima, sob diferentes situações, tamanho da amostra e herdabilidade. Essa autora pôde verificar que o número de graus de liberdade, definido como citado no parágrafo anterior, apresentou comportamentos adequados, resultando numa estatística de teste poderosa, com controle satisfatório do erro tipo I.

1.4 Grau médio de dominância

Segundo Mather e Jinks (1984), uma herança dissômica (dois alelos A e a) pode originar três genótipos AA, Aa e aa, sendo a contribuição dos homocigotos quantificada por $\pm a$ e do heterocigoto por d , em relação à média μ , conforme representado na Figura 1.

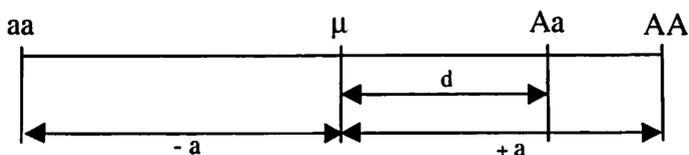


FIGURA 1. Valores genotípicos para um loco A

Considerando A o alelo que aumenta a expressão do caráter e utilizando esses desvios, podem-se avaliar os diferentes casos de interação alélica, ou seja, se

- (i) $d = 0$, não há dominância e a interação alélica é chamada aditiva.
- (ii) $d = a$, a interação alélica é de dominância completa.
- (iii) $0 < d < a$, a interação alélica é de dominância parcial.
- (iv) $d > a$, a interação alélica é considerada sobredominante.

Pode-se ainda salientar que a relação d/a mede o que se denomina grau de dominância de um gene, o qual dá idéia da interação alélica. Assim, se este valor é zero, a interação alélica é aditiva; quando é igual a 1,0 há dominância completa; se estiverem compreendidos entre zero e um ocorre dominância parcial e, para os valores superiores a 1,0, ocorre sobredominância (Ramalho et al., 2000). Esta idéia se aplica tanto a poligenes quanto a um gene principal.

1.5 Herdabilidade

A herdabilidade pode ser definida como a razão da variância herdável (variância genética aditiva) pela variância fenotípica. Como o melhoramento de uma população por seleção depende da variação gênica aditiva, o coeficiente de herdabilidade serve para orientação sobre a quantidade relativa de variação genética que é utilizável no melhoramento (Ramalho et al., 1993). Assim, quantifica-se o grau de facilidade ou dificuldade de obter sucesso por seleção em programas de melhoramento. Por esta razão, a herdabilidade entra em quase todas as fórmulas relacionadas com métodos de melhoramento e muitas das decisões práticas sobre os procedimentos a serem adotados dependem da sua magnitude.

Apesar das diferenças nas herdabilidades de uma certa característica, as estimativas existentes na literatura mostram que cada caráter apresenta uma amplitude de valores que lhe é peculiar (Ramalho et al., 2000).

2 Taxa de erros tipo I e II e poder do teste

Erros tipo I e II são os erros que podem ser cometidos ao se fazer um teste de hipótese. Segundo Mayer (1984), o ideal é manter uma probabilidade relativamente pequena de cometê-los.

Suas definições podem ser encontradas em Mood et al. (1974). O erro tipo I é definido como o erro que se comete ao rejeitar a hipótese nula, quando esta é verdadeira, com probabilidade representada por α e identificada como o nível de significância do teste. Já o erro tipo II é o erro que se comete ao aceitar a hipótese nula, quando na verdade esta é falsa, com probabilidade representada por β . O pesquisador pode controlar o erro tipo I decidindo o nível de risco α .

que ele está disposto a tolerar, o que não acontece com o tipo II, que não pode ser controlado diretamente.

Também encontrado em Mood et al. (1974), o poder do teste é definido como a probabilidade de rejeitar a hipótese nula, quando ela é falsa, ou seja, a probabilidade de se tomar a decisão correta, dada por $(1 - \beta)$.

Decisão estatística	Situação efetiva	
	H_0 verdadeiro	H_0 falso
Não rejeitar H_0	Confiança $(1 - \alpha)$	Erro Tipo II (β)
Rejeitar H_0	Erro Tipo I (α)	Poder $(1 - \beta)$

3 Técnicas de reamostragem

Tais técnicas, apesar de terem sido criadas há muitos anos, só se desenvolveram com o advento dos recursos computacionais, já que exigem grande quantidade de cálculos.

Técnicas de reamostragem são muito úteis em estudos de populações naturais, já que podem ser uma forma de contornar o problema da falta de repetições, além de, segundo Carlini-Garcia (2001), permitir a obtenção de estimativas de erros associados a essas estimativas, de intervalos de confiança para os parâmetros e testes de hipóteses e, em alguns casos, a distribuição dos parâmetros de interesse.

A seguir serão destacadas as principais técnicas de reamostragem e os intervalos de confiança construídos a partir delas.

3.1 Jackknife

A técnica jackknife foi introduzida por Quenouille em 1949 e o termo “jackknife”, por Tukey (1958), referindo-se aos canivetes de bolso com

múltiplas funções que, embora fáceis de carregar, raramente realizam alguma das funções de modo ideal. A técnica foi proposta com o intuito de descrever uma abordagem geral para testar hipóteses e construir intervalos de confiança quando outros métodos melhores não podem ser empregados com facilidade.

Segundo Efron e Tibshirani (1993), jackknife pode ser definido como uma ferramenta utilizada para se estimar, de maneira não paramétrica, algumas medidas de variabilidade, como variância e viés de uma estatística de interesse, como descrito a seguir.

Seja uma amostra (X_1, X_2, \dots, X_n) e um estimador $\hat{\theta} = s(x)$. Uma amostra jackknife é definida como $x_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, para $i = 1, 2, \dots, n$. Ou seja, a i -ésima amostra jackknife é amostra original sem a i -ésima observação. Logo, $\hat{\theta}_{(i)} = s(x_{(i)})$ é a i -ésima repetição jackknife de $\hat{\theta}$.

Assim, como em Efron e Tibshirani (1993), a estimativa jackknife do erro padrão é definida como

$$se(\hat{\theta}_{(.)}) = \left[\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]^{1/2},$$

em que $\hat{\theta}_{(.)} = \sum_{i=1}^n \frac{\hat{\theta}_{(i)}}{n}$ é o estimador jackknife de θ .

O intervalo de confiança aproximado e baseado na distribuição normal é então dado por:

$$IC: \hat{\theta} \pm z_{\alpha/2} se(\hat{\theta}_{(.)}),$$

em que $z_{\alpha/2}$ é o quantil $\alpha/2$ superior de distribuição normal padrão.

E o viés é dado por $(n-1)(\hat{\theta}_{(.)} - \hat{\theta})$.

Manly (1998) dá um outra forma de cálculo do erro padrão e conseqüentemente do viés, por meio do que ele chama de pseudo-valores, porém a idéia permanece a mesma.

3.2 Bootstrap

A metodologia bootstrap foi formalizada por Efron (1979). Esta técnica é baseada na reamostragem de dados reais para revelar algum padrão neles presente, assim como na randomização e jackknife; no entanto, neste caso a reamostragem é feita com reposição.

As duas principais vantagens das técnicas bootstrap em relação aos métodos mais tradicionais, segundo Hoyle e Cameron (2003), são: (a) eles são mais robustos, comportando-se melhor com dados que não concordam com as suposições dos padrões paramétricos e (b) eles são, em geral, simples de implementar, substituindo formas complexas com esforço computacional.

O procedimento consiste em obter amostras constituídas por n valores reamostrados com reposição a partir de um conjunto de dados reais observados de uma população (X_1, X_2, \dots, X_n) .

O processo é realizado inúmeras vezes, obtendo-se estimativas bootstrap dos parâmetros de interesse, as quais geram uma distribuição, denominada distribuição bootstrap, que possibilita realizar testes de hipótese e construir intervalos de confiança. Alguns exemplos de aplicação serão considerados a seguir.

Carlini-Garcia (2001) fez uso dos métodos de reamostragem bootstrap na estimação de parâmetros genéticos, utilizando dados reais, além de simulados em algumas situações, destacando a eficiência de tais métodos.

Carvalho (1996) aplicou a metodologia bootstrap como uma alternativa de análise de dados do sistema agroflorestral, os quais, por suas características

particulares, impossibilitam a aplicação de métodos paramétricos, em geral; o autor observou que o método bootstrap fornece estimativas bastante acuradas.

Os intervalos de confiança mais comuns, quando se utilizam técnicas bootstrap, são os ditos intervalos percentis. Segundo Manly (1998), o uso desse tipo de intervalo é uma tentativa de aproximar os percentis da distribuição de um estimador usando percentis gerados por bootstrap. O mais simples é o intervalo de confiança percentil simples; pode-se citar ainda sua versão melhorada, o método BCa (bootstrap com correção para o viés e aceleração), o qual corrige algumas deficiências do anterior. Estes intervalos são descritos a seguir, conforme Efron e Tibshirani (1993).

3.2.1 Intervalo de confiança percentil simples

Dada a distribuição bootstrap de um estimador, o intervalo de confiança percentil com coeficiente de confiança igual a $1 - 2\alpha$ é definido pelos percentis α e $1 - \alpha$ dessa distribuição; logo, os valores $\hat{\theta}^{(\alpha)}$ e $\hat{\theta}^{(1-\alpha)}$ representam os limites do intervalo considerado.

Segundo DiCiccio e Efron (1996), o intervalo percentil simples não apresenta propriedades de cobertura satisfatórias, além de apresentar viés; assim, o intervalo proposto a seguir, segundo os mesmos autores, seria uma melhor alternativa.

3.2.2 Intervalo de confiança BCa (correção para o viés acelerado)

O intervalo de confiança BCa também é obtido da distribuição bootstrap, de forma que o intervalo com coeficiente de confiança igual a $1 - 2\alpha$ é

BCa : $\left[\hat{\theta}^*(\alpha_1); \hat{\theta}^*(\alpha_2) \right]$, em que

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right) \text{ e } \alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right), \text{ e}$$

$z^{(\alpha)}$ é o α -ésimo ponto percentil da distribuição normal padronizada; e $\Phi(\cdot)$ indica a função de distribuição normal padrão acumulada.

Pode ser observado que o intervalo depende das estimativas de \hat{a} e \hat{z}_0 , denominados constante de aceleração e correção do viés, respectivamente, que são dadas por

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(i)})^3}{6 \left[\sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(i)})^2 \right]^{3/2}}, \text{ em que}$$

$$\hat{\theta}_{(i)} = \frac{\sum_{i=1}^n \hat{\theta}_{(i)}}{n} \text{ é a estimativa jackknife de } \theta;$$

$\hat{\theta}_{(i)}$ é uma estimativa de θ obtida a partir da amostra original sem a i -ésima observação;

n é o tamanho da amostra original;

E \hat{z}_0 é a proporção de estimativas menores que $\hat{\theta}^*$, ou seja,

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{\theta}_B^* < \hat{\theta}\}}{B}\right), \text{ em que}$$

$\hat{\theta}$: é uma estimativa de θ da amostra original;

$\hat{\theta}_B^*$: é uma estimativa de θ para a B -ésima repetição bootstrap;

B : número de repetições bootstrap;

Assim, de posse das estimativas de \hat{a} e \hat{z}_0 , determinam-se os percentis do intervalo de confiança BCa.



DiCiccio e Efron (1996) comentam que, em geral, esse intervalo requer cerca de 2000 amostras bootstrap.

Pereira et al. (2000) fizeram uso de intervalos de confiança “bootstrap” como ferramenta para classificar raças de nematóide de cisto da soja. Avaliando as estimativas “bootstrap” da média aritmética dos genótipos de 4 variedades de soja e os intervalos de confiança percentil simples e BCa, do parâmetro de diferenciação da cultivar Lee, pôde concluir que o intervalo de confiança percentil simples foi uma estimativa intervalar precisa para o parâmetro de diferenciação.

Hoyle e Cameron (2003) compararam intervalos de confiança bootstrap na estimação de parâmetros associados à pesca recreativa, considerando a cobertura do intervalo e o viés relativo. O intervalo que apresentou melhor cobertura foi o bootstrap-t; já o viés esteve presente em todos.

REFERÊNCIAS BIBLIOGRÁFICAS

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.

BESERRA Jr., J. E. A. *Herança da resistência ao Watermelon mosaic virus (WMV) em melancia (Citrullus lanatus) e sua relação com a resistência ao Papaya ringspot virus (PRSV-W)*. 2004. 65 p. Dissertação (Mestrado em Fitopatologia) – Universidade Federal de Lavras, Lavras, MG.

CARLINI-GARCIA, L. A. *Uso do método de reamostragem bootstrap na estimação de parâmetros genéticos populacionais*. 2001. 162 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba.

CARVALHO, A. C. A.; *O método bootstrap e sua aplicação em análise de dados agroflorestais com variáveis aleatórias do tipo razão*. 1996. 89 p. Dissertação (Mestrado em estatística e Experimentação Agropecuária) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba.

CHANGJIAN, J.; XUEBIAO, P.; MINGHONG, G. The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. *Genetics*, Baltimore, v. 136, n. 1, p. 383-394, Jan. 1994.

DICICCIO, T. J.; EFRON, B. Bootstrap Confidence Intervals. *Statistical Science*, Hayward, v. 11, n. 3, p. 189-228, Aug. 1996.

DRAPER, R. N.; SMITH H. *Applied regression analysis*. 3. ed. New York: John Wiley, 1998. 706 p.

EFRON, B. Bootstrap methods: another look at jackknife. *Annals of Statistics*, Hayward, v. 7, n. 1, p. 1-26, 1979.

EFRON, B.; TIBSHIRANI, R. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993. 436 p.

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.

GONÇALVES, R. P. **Avaliação da razão de verossimilhança generalizada em testes de hipótese sobre o controle genético de características contínuas.** 2004. 37 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

HOYLE, S. D.; CAMERON, D. S. Confidence intervals on catch estimates from a recreational fishing survey: a comparison of bootstrap methods. *Fisheries Management and Ecology*, Oxford, v. 10, n. 2, p. 97-108, Apr. 2003.

LIMA NETO, A. F. **Herança e resistência de genótipos de tomateiro a *Ralstonia solanacearum*, agente da murcha bacteriana.** 2001. 90 p. Dissertação (Mestrado em Fitopatologia) – Universidade Federal de Lavras, Lavras, MG.

LOISEL, P.; GOFFINET, B.; MONOD, H.; OCA, G. M. Detecting a major gene in an F₂ population. *Biometrics*, Raleigh, v. 50, n. 2, p. 512-516, Jun. 1994.

MANLY, B. F. J. **Randomization, bootstrap and Monte Carlo methods in biology.** 2. ed. London: Chapman-Hall, 1998. 399 p.

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

MAYER, P. L. **Probabilidade: aplicações à estatística.** Tradução do prof. Ruy de C. B. Lourenço Filho. 2. ed. Rio de Janeiro: LTC, 1984.

MENEZES, C. B. **Herança da partenocarpia e da resistência ao vírus da mancha anelar do mamoeiro – estirpe melancia (PRSV-W) em abobrinha (*Cucurbita pepo* L.).** 2003. 58 p. Tese (Doutorado em Fitotecnia) – Universidade Federal de Lavras, Lavras, MG.

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

OLIVEIRA, A. C. B. **Herança da resistência ao vírus da mancha anelar do mamoeiro-estirpe melancia (PRSV-W) em *Cucurbita moschata* duch. e sua introgressão em *Cucurbita pepo* L.** 1999. 74 p. Tese (Doutorado em Fitotecnia) – Universidade Federal de Lavras, Lavras, MG.

PEREIRA, J. E. SILVA, J. F. V. DIAS, W. F. SOUZA, G. S.; Intervalo de confiança “bootstrap” como ferramenta para classificar raças do nematóide de cisto da soja. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 2, p. 271-275, fev. 2000.

QUENOUILLE, M. Approximations tests of correlation in time series. **Journal of Royal Statistical Society B**, London, v. 11, n. 1, p. 18-84, 1949.

RAMALHO, M.; SANTOS, J. B.; PINTO, C. B. **Genética na agropecuária**. 7. ed. Lavras: Globo, 2000. 359 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. 271p.

RÊGO, E. R. do. **Herança da cor e caracterização de frutos de tomate mutante, normal, F_1 e F_2** . 1997. 66 p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.

SILVA, W. P. **Estimadores de máxima verossimilhança em misturas de densidades normais: uma aplicação em genética**. 2003. 60 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

SOUZA SOBRINHO, F. **Herança da reação de resistência à raça 2 de *Meloidogyne incognita* na pimenta *Capsicum annuum* L. cv Carolina Cayenne**. 1998. 57 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG.

TUKEY, J. W. Bias and confidence in nota quite large samples. **Annals of Mathematical Statistics**, Hayward, v. 29, n. 2, p. 614, June 1958.

CAPÍTULO 2

AVALIAÇÃO DA QUALIDADE DO TESTE DE RAZÃO DE VEROSSIMILHANÇA, QUANTO AO CONTROLE DO ERRO TIPO I E DO PODER DO TESTE

RESUMO

MOURA, Gisele Martins. Avaliação da qualidade do teste de razão de verossimilhança, quanto ao controle do erro tipo I e do poder do teste. In: ____. **Erro tipo I, poder do teste e intervalos de confiança com técnicas de reamostragem em modelos de herança genética utilizando misturas de normais**. Lavras: UFLA, 2005. 68p. (Dissertação - Mestrado em Agronomia/Estatística e Experimentação Agropecuária).¹

Em programas de melhoramento, se a característica em estudo é contínua, tem-se interesse em determinar se há ação de poligenes e/ou de gene principal. Uma das metodologias que fazem esta determinação simultaneamente é o modelo de herança utilizando misturas de normais, que utiliza o teste de razão de verossimilhança generalizada, considerando aproximação de qui-quadrado. No presente trabalho, foi avaliada a qualidade da estatística desse teste, quanto à taxa de erro tipo I e o poder do teste, utilizando simulação de dados, sob diferentes situações, comuns em tais estudos, variando: o tamanho da amostra, a herdabilidade e o grau médio de dominância, tanto para o gene principal quanto para os poligenes. O controle do erro tipo I, sob H_0 , foi avaliado em duas situações: na ausência do gene principal e na ausência de poligenes. Já o poder do teste foi avaliado com ambos presentes. Os resultados mostraram que o teste de razão de verossimilhança proporcionou um controle adequado do erro tipo I, porém rigoroso, já que, em todos os casos, a taxa de rejeição empírica de H_0 esteve abaixo do nível de significância nominal considerado (5%). O poder se mostrou elevado, tanto na detecção de poligenes quanto na do gene principal, com exceção da situação de detecção do gene principal quando a proporção da variação genotípica do mesmo era baixa. Observou-se, ainda, que os diferentes valores do grau médio de dominância não influenciaram o teste. Enfim, o teste de razão de verossimilhança utilizando o critério de decisão assintótico mostrou-se excelente em estudos de herança.

¹ **Comitê Orientador:** Dr. Eduardo Bearzoti – UFLA.(Orientador), Dr. Daniel Furtado Ferreira – UFLA

ABSTRACT

MOURA, Gisele Martins. Evaluation of the likelihood ratio test with regard type I error rate and the power of the test. In: _____. **Type I Error, Power Test and Confidence Intervals With Resampling Techniques in Genetic Models Using Mixtures of Normal Densities**. Lavras: UFLA, 2005. 68p. (Dissertation - Master in Agronomy/Major in Statistics and Agricultural Experimentation).¹

In plant breeding programs, if the trait under consideration is continuous, it is interesting to investigate whether there is the action of polygenes and/or of a major gene. One approach to carry out such investigations simultaneously is to use a genetic model that uses mixtures of normal densities and the likelihood ratio test considering the chi-square approximation. In this study the test statistic was evaluated with regard type I error rate and the power of the test, using simulated data. Different conditions were investigated, varying sample size, heritability and the degree of dominance of both polygenes and the major gene. Type I error rates were accounted for under two null hypotheses: without polygenes and without a major gene. The power of the test was evaluated with both kinds of genes present. Results showed that the likelihood ratio test was successful to control type I error rate, but it was somewhat conservative, since the rates were always below the nominal significance level (5%). The power of the test was high to detect polygenes and the major gene in most cases, but not to detect the major gene when it explained a small portion of phenotypic variation. The degree of dominance did not affect the performance of the test statistic. In general, it could be observed that the likelihood ratio test using its asymptotic distribution was an excellent criterion.

¹ **Guidance Committee:** Dr. Eduardo Bearzoti – UFLA (Adviser), Dr. Daniel Furtado Ferreira – UFLA

INTRODUÇÃO

Uma das indagações que o geneticista de plantas costuma fazer diante do estudo de características contínuas é se há ação de um gene principal e/ou poligenes, já que esta informação possibilita determinar a melhor estratégia de melhoramento e assim otimizar o processo.

Tais estudos são feitos, em geral, separadamente, isto é, sem considerar um modelo único na estimação de parâmetros associados a poligenes e ao gene principal.

Quanto aos poligenes, geralmente a inferência é feita estimando-se componentes de variância, utilizando análise de variância e componentes de média, pela metodologia do teste de escala conjunto (Mather e Jinks, 1984; Ramalho et al., 1993). Já para o gene principal, existem diferentes propostas na literatura; porém, a maioria delas tem em comum a construção de classes com as quais testes de aderência de qui-quadrado são feitos.

Changjian et al. (1994) e Silva (2003) apresentam um modelo de herança contendo parâmetros referentes a poligenes e um gene principal, simultaneamente. Esse modelo admite que haja distribuição normal associada a cada geração e permite a aplicação do método de máxima verossimilhança na estimação dos parâmetros. A partir deste modelo geral, a construção de submodelos com menos parâmetros, de particular interesse, é possível. Assim, testes de razão de verossimilhança são feitos para verificar a existência de poligenes e/ou de gene principal.

A estatística destes testes segue uma distribuição aproximada de qui-quadrado, o que leva a indagar quanto à sua qualidade.

Diante disso, o objetivo do presente trabalho foi avaliar a qualidade do teste de razão de verossimilhança para os modelos de herança genética contendo parâmetros referentes a poligenes e gene principal, quanto ao controle do erro tipo I e do poder do teste em diferentes condições.

MATERIAL E MÉTODOS

Para o desenvolvimento deste trabalho foi considerado o modelo de herança conforme descrito por Silva (2003), contendo parâmetros referentes a poligenes e um gene principal, simultaneamente, considerando dados das gerações P_1 , P_2 , F_1 , F_2 , RC_1 e RC_2 . Considerou-se que os efeitos genéticos, tanto do gene principal quanto dos poligenes, apresentavam dominância, o que leva a parâmetros adicionais no modelo, conforme descrito na seção anterior.

Silva (2003) admite que haja uma distribuição normal associada a cada genótipo do gene principal, com médias e variâncias específicas, sendo que estas últimas dependem da geração. Aos retrocruzamentos e à geração F_2 correspondem misturas de distribuições normais, obedecendo a proporções definidas conforme a frequência esperada de cada genótipo do gene principal, em cada geração. Além disso, a variância ambiental (σ^2) foi considerada igual em todas as gerações.

A qualidade do teste de razão de verossimilhança para esses modelos de herança, quanto ao controle do erro tipo I sob H_0 , foi avaliada em duas situações: na ausência do gene principal e na ausência dos poligenes. Já o poder do teste foi avaliado com ambos presentes.

Para tanto, foram simulados dados para as seis gerações, conforme as densidades apresentadas no capítulo anterior. Consideraram-se 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, que foram considerados tamanhos de referência, representados, em conjunto, como “n” foi definido como amostra de tamanho n. De maneira análoga, simularam-se tamanhos $2n$ e $n/2$. Além disso, foram admitidos valores de herdabilidade iguais a 0,2 e 0,7, os quais acredita-se serem valores representativos de um valor baixo e um valor alto. Estabeleceram-se também os valores de grau médio de dominância iguais a 0, $\frac{1}{2}$, 1 e 2, tanto para os

poligenes quanto para o gene principal. Por fim, sem perda de generalidade, admitiu-se que $\mu = 1000$ e $\sigma^2 = 1000$. A obtenção dos demais parâmetros é descrita a seguir.

1 Cálculo dos parâmetros para avaliação do erro tipo I sob H_0

Definindo-se G como o grau médio de dominância, sua expressão é dada por $G = d/a$, ou seja, $d = Ga$, com a e d como definidos na seção 1.4. Consideraram-se neste trabalho valores de G iguais 0, $\frac{1}{2}$, 1 e 2, que dão a idéia das diferentes interações alélicas, conforme apresentado na seção 1.4.

A variância aditiva V_A , a variância de dominância V_D e o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância S_{AD} são dados por $V_A = \sum_{i=1}^g \frac{a_i^2}{2}$, $V_D = \sum_{i=1}^g \frac{d_i^2}{4}$ e

$$S_{AD} = -\sum_{i=1}^g \frac{a_i d_i}{2}, \text{ respectivamente (Ramalho et al., 2000).}$$

Assim, como $d = Ga$, tem-se

$$V_D = \sum_{i=1}^g \frac{(Ga_i)^2}{4} \Rightarrow V_D = \frac{G^2}{2} \sum_{i=1}^g \frac{a_i^2}{2} \Rightarrow V_D = \frac{G^2}{2} V_A \quad e$$

$$S_{AD} = -\sum_{i=1}^g \frac{a_i Ga_i}{2} = -G \sum_{i=1}^g \frac{a_i^2}{2} \Rightarrow S_{AD} = -G V_A$$

1.1 Modelos sob comparação: 1 e 5.

Confrontando-se os modelos 1 e 5, está-se testando a hipótese da existência de um gene principal; para isso, atribuiu-se valor igual a zero aos parâmetros A e D , que correspondem aos efeitos aditivo e de dominância do gene principal, respectivamente.

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 variação total explicada pelos poligenes, com variância genética de referência em F_2 .

Assim, substituindo $V_D = \frac{G^2}{2} V_A$, $A = D = 0$ e $\sigma^2 = 1000$ em h_p^2 , tem-se

$$h_p^2 = \frac{V_A + \frac{G^2}{2} V_A}{V_A + \frac{G^2}{2} V_A + 1000} \Rightarrow V_A + \frac{G^2}{2} V_A - V_A h_p^2 - \frac{G^2}{2} V_A h_p^2 = 1000 h_p^2$$

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

$$V_A = \frac{2000 h_p^2}{2 + G^2 (1 - h_p^2)} \quad \text{e} \quad V_D = \frac{1000 G^2 h_p^2}{2 + G^2 (1 - h_p^2)}$$

Considerou-se o número de poligenes $g = 100$ e $a_1 = a_2 = L = a_{100} = a$.

Com isso, $V_A = \frac{100}{2} a^2 \Rightarrow a = \frac{\sqrt{2V_A}}{10}$. Seja $[a]$ a soma dos efeitos aditivos dos

poligenes. Sua expressão é dada por $[a] = \sum_{i=1}^g a_i$; assim

$[a] = 100a \Rightarrow [a] = 100 \frac{\sqrt{2V_A}}{10}$. Além disso, seja $[d]$ a soma dos efeitos de

dominância dos poligenes, que é $[d] = \sum_{i=1}^g d_i$; como $d = Ga$, tem-se

$$[d] = \sum_{i=1}^g Ga_i = G \sum_{i=1}^g a_i = G[a] \Rightarrow [d] = G10 \sqrt{2V_A}.$$

Variando-se os coeficientes de herdabilidade (h_p^2) e os valores do grau médio de dominância (G) , já definidos, nestas expressões gerais, obtiveram-se os valores dos parâmetros para cada situação, como detalhado a seguir.

1.1.1 Valores dos parâmetros para $G = 0$

(i) Sob $h_p^2 = 0,2$

$$A = D = [d] = V_D = S_{AD} = 0 ; V_A = 250 \text{ e } [a] = 223,607$$

(ii) Sob $h_p^2 = 0,7$

$$A = D = [d] = V_D = S_{AD} = 0 ; V_A = 2333,333 \text{ e } [a] = 683,130$$

1.1.2 Valores dos parâmetros para $G = 1/2$

(i) Sob $h_p^2 = 0,2$

$$A = D = 0 ; V_A = 222,222 ; V_D = 27,778 ; \\ S_{AD} = -111,111 ; [a] = 210,818 \text{ e } [d] = 105,409$$

(ii) Sob $h_p^2 = 0,7$

$$A = D = 0 ; V_A = 2074,074 ; V_D = 259,259 ; \\ S_{AD} = -1037,037 ; [a] = 644,061 \text{ e } [d] = 322,030$$

1.1.3 Valores dos parâmetros para $G = 1$

(i) Sob $h_p^2 = 0,7$

$$A = D = 0 ; V_A = 166,667 ; V_D = 83,335 ; \\ S_{AD} = -166,667 ; [a] = 182,576 \text{ e } [d] = 182,576$$

(ii) Sob $h_p^2 = 0,7$

$$A = D = 0 ; V_A = 1555,556 ; V_D = 777,778 ; \\ S_{AD} = -1555,556 ; [a] = 557,774 \text{ e } [d] = 557,774$$

1.1.4 Valores dos parâmetros para G = 2

(i) Sob $h_p^2 = 0,2$

$$A = D = 0 ; V_A = 83,333 ; V_D = 166,667 ;$$

$$S_{AD} = -166,667 ; [a] = 129,099 \text{ e } [d] = 258,198$$

(ii) Sob $h_p^2 = 0,7$

$$A = D = 0 ; V_A = 777,777 ; V_D = 1555,556 ;$$

$$S_{AD} = -1555,556 ; [a] = 394,405 \text{ e } [d] = 788,810$$

1.2 Modelos sob comparação: 1 e 7

Confrontando-se os modelos 1 e 7, está-se testando a hipótese da existência de poligenes, Para isso atribuiu-se valor igual a zero aos parâmetros [a], [d], V_A , V_D e S_{AD} , que correspondem à soma dos efeitos aditivos dos poligenes, à soma dos efeitos de dominância dos poligenes, às variâncias aditiva e de dominância e ao componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância, respectivamente.

Definindo-se h_M^2 como o coeficiente de herdabilidade do gene principal,

sua expressão é dada por
$$h_M^2 = \frac{\frac{A^2}{2} + \frac{D^2}{4}}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$$
, ou seja, é a fração da

variação total explicada pelo gene principal, com variância genética de referência em F_2 .

Supondo-se $D = GA$ e substituindo-se $[a] = [d] = V_A = V_D = S_{AD} = 0$ e $\sigma^2 = 1000$ em h_M^2 , obtém-se

$$h_M^2 = \frac{\frac{A^2}{2} + \frac{(GA)^2}{4}}{\frac{A^2}{2} + \frac{(GA)^2}{4} + 1000} \Rightarrow 2A^2 + G^2A^2 - h_M^2 2A^2 - h_M^2 G^2A^2 = 4000h_M^2$$

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

$$\text{Assim, } D = 2G \sqrt{\frac{1000h_M^2}{(2+G^2)(1-h_M^2)}}$$

Variando-se os coeficientes de herdabilidade (h_M^2) e os valores do grau médio de dominância (G), já definidos, nestas expressões gerais, obtiveram-se os valores dos parâmetros para cada situação, como detalhado a seguir.

1.2.1 Valores dos parâmetros para G = 0

(i) Sob $h_M^2 = 0,2$

$$[a] = [d] = V_A = V_D = S_{AD} = D = 0 ; A = 22,361$$

(ii) Sob $h_M^2 = 0,7$

$$[a] = [d] = V_A = V_D = S_{AD} = D = 0 ; A = 68,313$$

1.2.2 Valores dos parâmetros para G = 1/2

(i) Sob $h_M^2 = 0,2$

$$[a] = [d] = V_A = V_D = S_{AD} = 0 ; A = 21,082 ; D = 10,541$$

(ii) Sob $h_M^2 = 0,7$

$$[a] = [d] = V_A = V_D = S_{AD} = 0 ; A = 64,406 ; D = 32,203$$

1.2.3 Valores dos parâmetros para G = 1

(i) Sob $h_M^2 = 0,2$

$$[a] = [d] = V_A = V_D = S_{AD} = 0 ; A = D = 18,257$$

(ii) Sob $h_M^2 = 0,7$

$$[a] = [d] = V_A = V_D = S_{AD} = 0 ; A = D = 55,777$$

1.2.4 Valores dos parâmetros para $G = 2$

(i) Sob $h_M^2 = 0,2$

$$[a] = [d] = V_A = V_D = S_{AD} = 0 ; A = 12,909 ; D = 25,819$$

(ii) Sob $h_M^2 = 0,7$

$$[a] = [d] = V_A = V_D = S_{AD} = 0 ; A = 39,441 ; D = 78,881$$

2 Cálculo dos parâmetros para avaliação do poder do teste

Para o desenvolvimento de tal avaliação, estabeleceu-se a partição $h^2 = h_M^2 + h_p^2$, sendo h_M^2 o coeficiente de herdabilidade do gene principal e h_p^2 o coeficiente de herdabilidade dos poligenes.

Como já foi definido, tem-se

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

Substituindo-se $D = GA$ e $V_D = \frac{G^2}{2}V_A$ nas expressões de h_p^2 e h_M^2 , obtém-se

$$h_p^2 = \frac{V_A + \frac{G^2}{2}V_A}{\frac{A^2}{2} + \frac{G^2A^2}{4} + V_A + \frac{G^2}{2}V_A + \sigma^2} \quad \text{e} \quad h_M^2 = \frac{\frac{A^2}{2} + \frac{G^2A^2}{4}}{\frac{A^2}{2} + \frac{G^2A^2}{4} + V_A + \frac{G^2}{2}V_A + \sigma^2}$$

Com algumas manipulações, isola-se V_A na expressão de h_p^2 :

$$V_A = \frac{(2+G^2)A^2h_p^2 + 4\sigma^2h_p^2}{(4+G^2)(1-h_p^2)}$$

Agora, substituindo-se V_A na expressão de h_M^2 acima e com algumas manipulações e cancelamentos, tem-se explicitado o valor de A (valor médio do gene principal):

$$A = 2 \sqrt{\frac{1000h_M^2}{(2+G^2)(1-(h_p^2+h_M^2))}}$$

e como $D = GA$, tem-se $D = 2G \sqrt{\frac{1000h_M^2}{(2+G^2)(1-(h_p^2+h_M^2))}}$

Agora, substituindo-se esta expressão de A em V_A acima, tem-se

$$V_A = \frac{\frac{4\sigma^2h_M^2h_p^2}{(1-(h_p^2+h_M^2))} + 4\sigma^2h_p^2}{(4+2G^2)(1-h_p^2)}, \text{ como } V_D = \frac{G^2}{2} V_A, \text{ tem-se}$$

$$V_D = \frac{G^2}{2} \frac{\frac{4\sigma^2h_M^2h_p^2}{(1-(h_p^2+h_M^2))} + 4\sigma^2h_p^2}{(4+2G^2)(1-h_p^2)}.$$

E, por fim, analogamente ao item 3.1, tem-se $S_{AD} = -G V_A$.

Considerando-se as possíveis combinações dos coeficientes de herdabilidade (h_M^2 e h_p^2) e os valores do grau médio de dominância (G), já definidos, obtiveram-se os valores dos parâmetros para cada situação, como detalhado a seguir.

2.1 Valores dos parâmetros para $G = 0$

(i) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,2$

$$D = [d] = V_D = S_{AD} = 0 ;$$

$$A = 25,819 ; [a] = 258,198 ; V_A = 333,333$$

(ii) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,7$
 $D = [d] = V_D = S_{AD} = 0$;
 $A = 118,321$; $[a] = 632,455$; $V_A = 2000$

(iii) Sob $h_p^2 = 0,7$ e $h_M^2 = 0,2$
 $D = [d] = V_D = S_{AD} = 0$;
 $A = 63,245$; $[a] = 1183,215$; $V_A = 7000$

2.2 Valores dos parâmetros para $G = 1/2$

(i) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,2$
 $A = 24,343$; $D = 12,172$; $[a] = 243,432$; $[d] = 121,716$
 $V_A = 296,296$; $V_D = 37,037$; $S_{AD} = -148,148$

(ii) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,7$
 $A = 111,554$; $D = 55,777$; $[a] = 596,284$; $[d] = 298,142$
 $V_A = 1777,777$; $V_D = 222,222$; $S_{AD} = -888,888$

(iii) Sob $h_p^2 = 0,7$ e $h_M^2 = 0,2$
 $A = 59,628$; $D = 29,814$; $[a] = 1115,5467$; $[d] = 557,773$
 $V_A = 6222,222$; $V_D = 777,777$; $S_{AD} = -3111,111$

2.3 Valores dos parâmetros para $G = 1$

(i) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,2$
 $A = D = 21,082$; $[a] = [d] = 210,817$
 $V_A = 222,222$; $V_D = 111,111$; $S_{AD} = -222,222$

(ii) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,7$
 $A = D = 96,609$; $[a] = [d] = 516,398$
 $V_A = 1333,334$; $V_D = 666,667$; $S_{AD} = -1333,334$

(iii) Sob $h_p^2 = 0,7$ e $h_M^2 = 0,2$

$$A = D = 51,639 \quad ; \quad [a] = [d] = 966,092$$

$$V_A = 4666,666 \quad ; \quad V_D = 2333,333 \quad ; \quad S_{AD} = -4333,333$$

2.4 Valores dos parâmetros para $G = 2$

(iv) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,2$

$$A = 14,907 \quad ; \quad D = 29,814 \quad ; \quad [a] = 149,071 \quad ; \quad [d] = 298,142$$

$$V_A = 111,111 \quad ; \quad V_D = 222,222 \quad ; \quad S_{AD} = -222,222$$

(v) Sob $h_p^2 = 0,2$ e $h_M^2 = 0,7$

$$A = 68,313 \quad ; \quad D = 136,626 \quad ; \quad [a] = 365,148 \quad ; \quad [d] = 730,296$$

$$V_A = 666,666 \quad ; \quad V_D = 1333,333 \quad ; \quad S_{AD} = -1333,333$$

(vi) Sob $h_p^2 = 0,7$ e $h_M^2 = 0,2$

$$A = 36,515 \quad ; \quad D = 73,029 \quad ; \quad [a] = 683,130 \quad ; \quad [d] = 1366,260$$

$$V_A = 2333,333 \quad ; \quad V_D = 4666,666 \quad ; \quad S_{AD} = -4666,666$$

Especificados os 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 (F_2 , RC_1 e RC_2), em uma primeira etapa, sorteava-se a qual componente da mistura o indivíduo pertencia, conforme probabilidades apropriadas, obedecendo às devidas proporções 1:1 e 1:2:1, nos retrocruzamentos e na F_2 , respectivamente. Em seguida, sorteava-se uma realização da variável aleatória normal, correspondente ao componente em questão, pelas técnicas usuais de simulação.

As rotinas computacionais necessárias foram escritas em Kylix (versão Delphi para Linux).

Os modelos foram ajustados como em Silva (2003), utilizando o método numérico iterativo que combina os métodos Quase-Newton e Powell. Para tanto

foi especificado o número máximo de 500 iterações. Assim, nem sempre todas as amostras simuladas eram aproveitadas devido ao critério de convergência não ser satisfeito, ou seja, quando eventualmente o limite máximo de 500 iterações por amostra era atingido e a convergência não acontecia. As amostras que atingiam a convergência foram chamadas de amostras válidas.

Com base no número de amostras válidas apresentadas por Gonçalves (2004), estimou-se aproximadamente o número de amostras simuladas necessárias para que o número de amostras válidas fosse igual ou superior a 1000. Assim, sempre que este valor não era atingido, a simulação era repetida, aumentando-se o número de amostras. Ressaltando que, em algumas poucas exceções, o número de iterações foi aumentado para 1000.

Foi obtida a taxa de rejeição empírica de H_0 em cada configuração, que é a proporção de amostras válidas que foram rejeitadas. A partir daí, foi construído um intervalo de confiança exato para a mesma, baseado na distribuição F, para confrontar com o valor nominal de significância utilizado (5%).

RESULTADOS E DISCUSSÃO

1 Avaliação do erro tipo I sob H_0

Nas Tabelas 1 a 12 são apresentados os resultados obtidos quanto ao controle do erro tipo I, sob as diferentes condições mencionadas. As Tabelas de 1 a 4 apresentam os resultados com amostra de tamanho $n/2$, grau médio de dominância 0, $1/2$, 1 e 2, e coeficientes de herdabilidade 0,2 e 0,7. Já as Tabelas de 5 a 8 e as Tabelas de 9 a 12 mostram os resultados com as mesmas condições exceto o tamanho da amostra, que foram considerados n e $2n$, respectivamente.

TABELA 1 Resultados sob H_0 , com amostra de tamanho $n/2$, grau médio de dominância igual a 0 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	10000	1779	0,0067453	[0,0034900; 0,0117529]
1 versus 5	$A = D = 0$	0,7	13000	1939	0,0201134	[0,0143407; 0,0273939]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	2800	1555	0,0083601	[0,0044586; 0,0142534]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	2700	1052	0,0209125	[0,0131509; 0,0314912]

TABELA 2 Resultados sob H_0 , com amostra de tamanho $n/2$, grau médio de dominância igual a 1/2 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	9000	1698	0,0064782	[0,0032381; 0,0115614]
1 versus 5	$A = D = 0$	0,7	8000	762	0,0236220	[0,0140584; 0,0370764]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	2300	1461	0,0095824	[0,0052483; 0,0160253]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	2600	1891	0,0169222	[0,0116028; 0,0238057]

TABELA 3 Resultados sob H_0 , com amostra de tamanho $n/2$, grau médio de dominância igual a 1 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	8500	1611	0,0080695	[0,0043034; 0,0137594]
1 versus 5	$A = D = 0$	0,7	8000	386	0,0259067	[0,0124911; 0,0471247]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	1800	1258	0,0127186	[0,0072866; 0,0205720]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	1500	1263	0,0213776	[0,0141340; 0,0309517]

TABELA 4 Resultados sob H_0 , com amostra de tamanho $n/2$, grau médio de dominância igual a 2 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	7000	1315	0,0136882	[0,0081320; 0,0215471]
1 versus 5	$A = D = 0$	0,7	7000	532	0,0357142	[0,0216365; 0,0552113]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	1800	1268	0,0070977	[0,0032504; 0,0134305]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	1500	1107	0,0180668	[0,0110696; 0,0277652]

TABELA 5 Resultados sob H_0 , com amostra de tamanho n , grau médio de dominância igual a 0 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	10000	1004	0,0079681	[0,0034459; 0,0156391]
1 versus 5	$A = D = 0$	0,7	15000	1121	0,0107047	[0,0055430; 0,0186239]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	2500	1009	0,0019821	[0,0002401; 0,0071417]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	3000	1032	0,0193798	[0,0118768; 0,0297723]

TABELA 6 Resultados sob H_0 , com amostra de tamanho n , grau médio de dominância igual a 1/2 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	11000	1124	0,0071174	[0,0030775; 0,0139752]
1 versus 5	$A = D = 0$	0,7	20000	697	0,0243902	[0,0142707; 0,0387641]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	2500	1361	0,0036737	[0,0011937; 0,0085516]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	3000	1993	0,0170597	[0,0118425; 0,0237582]

TABELA 7 Resultados sob H_0 , com amostra de tamanho n , grau médio de dominância igual a 1 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	11000	1052	0,0066539	[0,0026791; 0,0136608]
1 versus 5	$A = D = 0$	0,7	20000	246	0,0325203	[0,0141416; 0,0630660]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	2500	1232	0,0024271	[0,0005008; 0,0070767]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	3000	2232	0,0165770	[0,0116979; 0,0227774]

TABELA 8 Resultados sob H_0 , com amostra de tamanho n , grau médio de dominância igual a 2 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	11000	1010	0,0079207	[0,0034254; 0,0155466]
1 versus 5	$A = D = 0$	0,7	20000	376	0,0452127	[0,0265550; 0,0714036]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	2500	1536	0,0091145	[0,0049916; 0,0152452]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	3000	2024	0,0212450	[0,0154167; 0,0285111]

TABELA 9 Resultados sob H_0 , com amostra de tamanho $2n$, grau médio de dominância igual a 0 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	14000	791	0,0075853	[0,0027883; 0,0164354]
1 versus 5	$A = D = 0$	0,7	16000	741	0,0040485	[0,0008357; 0,0117857]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	4000	1171	0,0111016	[0,0059231; 0,0189090]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	3500	1042	0,0095969	[0,0046112; 0,0175776]

TABELA 10 Resultados sob H_0 , com amostra de tamanho $2n$, grau médio de dominância igual a $1/2$ e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	13500	695	0,0071942	[0,0023396; 0,0167071]
1 versus 5	$A = D = 0$	0,7	15000	179	0,0167597	[0,0034697; 0,0481956]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	4000	1765	0,0073654	[0,0039273; 0,0125620]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	3500	1932	0,0160455	[0,0109275; 0,0226985]

TABELA 11 Resultados sob H_0 , com amostra de tamanho $2n$, grau médio de dominância igual a 1 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	13000	502	0,0039840	[0,0004828; 0,0143171]
1 versus 5	$A = D = 0$	0,7	14000	26	0,0000000	[0,0000000; 0,1088304]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	3000	1367	0,0109729	[0,0061540; 0,0180334]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	2000	1123	0,0151380	[0,0088424; 0,0241268]

TABELA 12 Resultados sob H_0 , com amostra de tamanho $2n$, grau médio de dominância igual a 2 e diferentes valores de h^2

Modelos sob comparação	Hipótese de nulidade	h^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	12500	542	0,0110701	[0,0040726; 0,0239367]
1 versus 5	$A = D = 0$	0,7	13000	62	0,0161290	[0,0004082; 0,0866202]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	3000	1371	0,0072939	[0,0035030; 0,0133724]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	2000	1085	0,0258064	[0,0172150; 0,0370825]

Observa-se um número diferente de amostras válidas em cada configuração, já que, como mencionado, é um processo iterativo que tem um critério de convergência a ser satisfeito. Apesar de repetir e até mesmo aumentar o número de iterações, na tentativa de atingir as 1000 amostras válidas, em alguns casos isso não foi possível, devido ao elevado tempo computacional. Mas, acredita-se que a pouca convergência, nesses casos, se deve a um problema numérico, ou seja, tal configuração apresentou uma maior dificuldade de convergência. Por exemplo, o caso em que se está confrontando os modelos 1 e 5, com coeficiente de herdabilidade 0,7, na Tabela 11, em que, de 14000 amostras simuladas apenas 26 convergiram, o que levou a taxa de rejeição a ser zero.

A baixa proporção de amostras válidas em alguns casos pode dar abertura para o questionamento de um possível viés. Porém, pode-se citar o caso em que se está confrontando os modelos 1 e 7 com coeficiente de herdabilidade 0,7, na Tabela 3, em que de 1500 amostras simuladas, 1263 convergiram, cerca de 84% e, ainda assim, a taxa de rejeição se manteve baixa (0,0213776). E mais, em uma situação real, como se trata de uma amostra apenas, pode-se trabalhar com um número muito maior de iterações. Assim, acredita-se que, caso como o da Tabela 11, não apresentaria tais valores.

As taxas de erro tipo I para o teste de razão de verossimilhança desses modelos, em quase todas as configurações, foram significativamente menores do que o valor nominal (0,05), com exceção dos casos que se esta confrontando os modelos 1 e 5, com coeficiente de herdabilidade 0,7 nas Tabelas 4, 7, 8 e 12, nas quais os intervalos de confiança abrangem o valor nominal. No entanto, tais configurações têm em comum a baixa proporção de amostras válidas, o que pode ser uma justificativa para o fato.

De modo geral, o coeficiente de herdabilidade igual a 0,2 apresentou menores taxas de rejeição de H_0 e, quando era igual a 0,7, as taxas tenderam a

ser mais próximas ao valor nominal. Diante disso, seria interessante fazer tais simulações considerando outros valores para o coeficiente de herdabilidade, para tentar confirmar tal tendência.

Houve uma tendência geral, ainda que modesta, de, com o aumento do tamanho da amostra, a taxa de rejeição de H_0 aproxima-ser do nível de significância nominal, concordando com o esperado, já que trata-se de uma distribuição assintótica de qui-quadrado.

Os resultados de simulação mostraram, ainda, que os diferentes valores do grau médio de dominância não apresentaram diferença significativa no valor da taxa de rejeição empírica.

Enfim, esses resultados juntos permitiram concluir que o teste de razão de verossimilhança para a detecção de poligenes e/ou gene principal proporcionou um controle adequado do erro tipo I, porém, considerado rigoroso, principalmente nos casos com menor precisão, menores tamanho de amostra e coeficiente de herdabilidade. É importante destacar, ainda, que o fato de as taxas terem ficado abaixo do valor nominal (0,05) não significa um problema; ao contrário, dá indícios de que tal teste seria uma ótima alternativa. No entanto, a confirmação desse fato requer a análise do poder.

Esses resultados concordam com os apresentados por Gonçalves (2004). Contudo, neste trabalho, as taxas foram calculadas com base em um número, em geral, maior ou igual a 1000 amostras válidas, o que aumenta a precisão com que as taxas de rejeição são estimadas.

2 Avaliação do poder do teste

As Tabelas de 13 a 24 ilustram os resultados obtidos na avaliação do teste quanto ao poder, sob diferentes condições. As Tabelas 13 a 16 apresentam os resultados com amostra de tamanho $n/2$, grau médio de dominância 0, $\frac{1}{2}$, 1 e

2, e uma combinação dos coeficientes de herdabilidade h_M^2 e h_P^2 , iguais a 0,2 e 0,7. Já as Tabelas 17 a 20 e as Tabelas 21 a 24 mostram os resultados com as mesmas condições, exceto exceto o tamanho da amostra, que foram considerados n e $2n$, respectivamente. Neste caso, o nível de significância também foi 5%.

TABELA 13 Avaliação do poder do teste para amostra de tamanho $n/2$ e grau médio de dominância igual a 0

Modelos sob comparação	Hipótese de nulidade	h_P^2	h_M^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	4000	1592	0,0251256	[0,0180093; 0,0340579]
1 versus 5	$A = D = 0$	0,2	0,7	4500	1232	0,9553571	[0,9422850; 0,9661947]
1 versus 5	$A = D = 0$	0,7	0,2	8000	1544	0,0751295	[0,0624752; 0,0894263]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2000	1275	1,0000000	[0,9976531; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	2500	1092	1,0000000	[0,9972604; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	2500	1124	1,0000000	[0,9973383; 1,0000000]

TABELA 14 Avaliação do poder do teste para amostra de tamanho $n/2$ e grau médio de dominância igual a 1/2

Modelos sob comparação	Hipótese de nulidade	h_P^2	h_M^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	2100	1033	0,0416263	[0,0302861; 0,0556631]
1 versus 5	$A = D = 0$	0,2	0,7	2000	1065	0,9990610	[0,9947796; 0,9999762]
1 versus 5	$A = D = 0$	0,7	0,2	8000	1211	0,0974401	[0,0813179; 0,1155463]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	1800	1301	1,0000000	[0,9977000; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	1500	1015	1,0000000	[0,9970528; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	5000	1338	1,0000000	[0,9977635; 1,0000000]

TABELA 15 Avaliação do poder do teste para amostra de tamanho $n/2$ e grau médio de dominância igual a 1

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	2100	1090	0,0513761	[0,0390389; 0,0662005]
1 versus 5	$A = D = 0$	0,2	0,7	8500	994	1,0000000	[0,9969907; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	6000	847	0,0956316	[0,0772947; 0,1174758]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	1800	1284	1,0000000	[0,9976695; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	2500	1400	1,0000000	[0,9978624; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	7000	373	1,0000000	[0,9920007; 1,0000000]

TABELA 16 Avaliação do poder do teste para amostra de tamanho $n/2$ e grau médio de dominância igual a 2

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	2000	1102	0,0453720	[0,0338604; 0,0593825]
1 versus 5	$A = D = 0$	0,2	0,7	1500	966	0,9989648	[0,9942459; 0,9999737]
1 versus 5	$A = D = 0$	0,7	0,2	10000	1480	0,0837837	[0,0701673; 0,0990774]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	1500	1103	1,0000000	[0,9972877; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	1500	1115	1,0000000	[0,9973168; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	6500	1267	1,0000000	[0,9976383; 1,0000000]

TABELA 17 Avaliação do poder do teste para amostra de tamanho n e grau médio de dominância igual a 0

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	4500	1268	0,0299684	[0,0212928; 0,0409034]
1 versus 5	$A = D = 0$	0,2	0,7	5000	1014	1,0000000	[0,9970499; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	9000	1154	0,0649913	[0,0514594; 0,0807889]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2500	1222	1,0000000	[0,9975515; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	4000	1539	1,0000000	[0,9980553; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	9000	3596	1,0000000	[0,9991672; 1,0000000]

TABELA 18 Avaliação do poder do teste para amostra de tamanho n e grau médio de dominância igual a 1/2

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	mostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	2300	1039	0,0433108	[0,0317636; 0,0575265]
1 versus 5	$A = D = 0$	0,2	0,7	5000	2443	1,0000000	[0,9987745; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	8500	927	0,1154261	[0,0955646; 0,1377672]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2300	1572	1,0000000	[0,9980961; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	4000	2749	1,0000000	[0,9989108; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	8500	1822	1,0000000	[0,9983571; 1,0000000]

TABELA 19 Avaliação do poder do teste para amostra de tamanho n e grau médio de dominância igual a 1

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	2500	1258	0,0604133	[0,0478906; 0,0750358]
1 versus 5	$A = D = 0$	0,2	0,7	2800	1263	1,0000000	[0,9976308; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	6000	917	0,1145038	[0,0946134; 0,1369076]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2000	1314	1,0000000	[0,9977227; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	3500	1669	1,0000000	[0,9982066; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	6000	1034	1,0000000	[0,9971069; 1,0000000]

TABELA 20 Avaliação do poder do teste para amostra de tamanho n e grau médio de dominância igual a 2

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	3000	1544	0,0576424	[0,0465440; 0,0704579]
1 versus 5	$A = D = 0$	0,2	0,7	4000	2381	1,0000000	[0,9987426; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	12000	1166	0,1149228	[0,0971771; 0,1346362]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	1800	1183	1,0000000	[0,9974708; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	3500	2401	1,0000000	[0,9987530; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	7000	921	1,0000000	[0,9967525; 1,0000000]

TABELA 21 Avaliação do poder do teste para amostra de tamanho $2n$ e grau médio de dominância igual a 0

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	9000	1091	0,0320806	[0,0224449; 0,0443356]
1 versus 5	$A = D = 0$	0,2	0,7	5500	1193	1,0000000	[0,9974920; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	11000	1676	0,0525059	[0,0423211; 0,0642912]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2300	1104	1,0000000	[0,9972901; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	2800	1035	1,0000000	[0,9971097; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	2800	1017	1,0000000	[0,9970586; 1,0000000]

TABELA 22 Avaliação do poder do teste para amostra de tamanho $2n$ e grau médio de dominância igual a $1/2$

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	4500	1698	0,0747938	[0,0627306; 0,0883499]
1 versus 5	$A = D = 0$	0,2	0,7	3000	1748	1,0000000	[0,9982876; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	3000	837	0,1237113	[0,1025968; 0,1474019]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2300	1421	1,0000000	[0,9978940; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	2800	2099	1,0000000	[0,9985738; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	6000	942	1,0000000	[0,9968248; 1,0000000]

TABELA 23 Avaliação do poder do teste para amostra de tamanho 2n e grau médio de dominância igual a 1

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	2800	1249	0,1208967	[0,1033252; 0,1402739]
1 versus 5	$A = D = 0$	0,2	0,7	3500	1679	1,0000000	[0,9982173; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	7000	1100	0,1809090	[0,1668923; 0,2036852]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2000	1158	1,0000000	[0,9974163; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	6000	2830	1,0000000	[0,9989420; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	7000	1141	1,0000000	[0,9973779; 1,0000000]

TABELA 24 Avaliação do poder do teste para amostra de tamanho 2n e grau médio de dominância igual a 2

Modelos sob comparação	Hipótese de nulidade	h_p^2	h_m^2	Nº de amostras simuladas	Amostras válidas	Taxa de rejeição empírica	Intervalo de confiança
1 versus 5	$A = D = 0$	0,2	0,2	2500	1187	0,0918281	[0,0760057; 0,1097051]
1 versus 5	$A = D = 0$	0,2	0,7	2500	1868	1,0000000	[0,9983984; 1,0000000]
1 versus 5	$A = D = 0$	0,7	0,2	10000	2315	0,1269978	[0,1136965; 0,1412507]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,2	2300	1307	1,0000000	[0,9977105; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,2	0,7	2500	2012	1,0000000	[0,9985121; 1,0000000]
1 versus 7	$[a] = [d] = V_A = V_D = S_{AD} = 0$	0,7	0,2	9000	789	1,0000000	[0,9962103; 1,0000000]

Também neste caso observa-se um número diferente de amostras válidas em cada configuração. Também aqui, o trabalho de Gonçalves (2004) proporcionou uma estimativa aproximada do número de amostras necessárias para que o número de amostras válidas fosse igual ou superior a 1000. Porém, em algumas configurações isso não foi possível, devido às mesmas justificativas dadas na avaliação do erro tipo I.

Pode ser observado, nas Tabelas 13 a 24 um alto valor para o poder da estatística de teste de razão de verossimilhança para detecção de poligenes e/ou gene principal. A exceção foram os casos de detecção do gene principal, com coeficiente de herdabilidade referente a ele igual a 0,2, ou seja, muito baixo, o que faz com que tal gene assuma mais um caráter de poligene do que propriamente de gene principal. Assim, é natural que o poder para detectá-lo seja pequeno. Tal fato fica bem evidente se observar o caso de detecção do gene principal, com herdabilidade devido a ele igual a 0,7, já que, neste caso, o poder foi de 100%. Estes resultados foram concordantes com os de Gonçalves (2004).

Também neste caso, os diferentes valores de grau médio de dominância não tiveram influência significativa.

Em geral, pode-se dizer que a estatística de teste de razão de verossimilhança para detecção de poligenes e/ou gene principal, apesar de utilizar um critério de decisão aproximado, foi um critério excelente, já que apresentou um controle satisfatório do erro tipo I, porém, conservador, associado a um alto poder. Tais condições são ideais em um teste, pois, caso exista uma dessas classes de genes, com um poder de aproximadamente 100%, ela é detectada e com uma margem de erro ainda menor do que a que se dispôs cometer, o que dá segurança ao melhorista na aplicação do teste nas condições avaliadas.

Verificou-se, ainda, um baixo número de amostras válidas, conseqüência do número restrito de, no máximo, 500 iterações, considerado devido ao elevado

REFERÊNCIAS BIBLIOGRÁFICAS

CHANGJIAN, J.; XUEBIAO, P.; MINGHONG, G. The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. *Genetics*, Baltimore, v. 136, n. 1, p. 383-394, Jan. 1994.

GONÇALVES, R. P. **Avaliação da razão de verossimilhança generalizada em testes de hipótese sobre o controle genético de características contínuas.** 2004. 37 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

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

RAMALHO, M.; SANTOS, J. B.; PINTO, C. B. **Genética na agropecuária.** 7. ed. Lavras: Globo, 2000. 359 p.

SILVA, W. P. **Estimadores de máxima verossimilhança em misturas de densidades normais: uma aplicação em genética.** 2003. 60 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

CONCLUSÕES

Os resultados de simulações permitiram concluir que o teste de razão de verossimilhança para a detecção de poligenes e/ou gene principal em modelos de herança proporcional proporciona um controle adequado do erro tipo I, porém, um tanto conservador, já que apresentaram, em todos os casos, taxas de rejeição empírica de H_0 inferiores ao valor nominal proposto (0,05).

O gene principal é detectado com alto poder, exceto quando a proporção de variação genética explicada por ele era baixa. Já os poligenes apresentam um alto poder de detecção em todos os casos.

Há evidências de que, aumentando-se o número de amostras, a taxa de rejeição de H_0 se aproxima do valor nominal.

Em geral, menores valores do coeficiente de herdabilidade apresentaram menor taxa de rejeição de H_0 .

Tanto na avaliação do erro tipo I quanto do poder do teste para poligenes e gene principal, os diferentes valores do grau médio de dominância não apresentaram diferença significativa.

O teste de razão de verossimilhança, apesar de utilizar um critério de decisão aproximado, mostrou-se um excelente critério nos estudos de herança genética.

REFERÊNCIAS BIBLIOGRÁFICAS

CHANGJIAN, J.; XUEBIAO, P.; MINGHONG, G. The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. *Genetics*, Baltimore, v. 136, n. 1, p. 383-394, Jan. 1994.

GONÇALVES, R. P. **Avaliação da razão de verossimilhança generalizada em testes de hipótese sobre o controle genético de características contínuas.** 2004. 37 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

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

RAMALHO, M.; SANTOS, J. B.; PINTO, C. B. **Genética na agropecuária.** 7. ed. Lavras: Globo, 2000. 359 p.

SILVA, W. P. **Estimadores de máxima verossimilhança em misturas de densidades normais: uma aplicação em genética.** 2003. 60 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

CAPÍTULO 3

TÉCNICAS DE REAMOSTRAGEM JACKKNIFE E BOOTSTRAP NA ESTIMAÇÃO POR INTERVALO DOS PARÂMETROS GENÉTICOS

RESUMO

MOURA, Gisele Martins. Técnicas de reamostragem jackknife e bootstrap na estimação por intervalo dos parâmetros genéticos. In:____. **Erro tipo I, poder do teste e intervalos de confiança com técnicas de reamostragem em modelos de herança genética utilizando misturas de normais.** Lavras: UFLA, 2005. 68p. (Dissertação - Mestrado em Agronomia/Estatística e Experimentação Agropecuária).¹

No melhoramento genético, a estimativa intervalar é também muito importante, já que dá uma idéia da precisão do parâmetro, possibilitando ter uma idéia do erro que se está cometendo e até de se realizar testes. Esse trabalho teve como objetivo obter tais estimativas intervalares de parâmetros genéticos relativos a poligenes e a gene principal, utilizando técnicas de reamostragem jackknife e bootstrap, tendo, no bootstrap, sido calculados o intervalo percentil simples e BCa (correção para o viés e aceleração). Pôde-se concluir que os intervalos são uma boa alternativa para tais estimativas. Eles não apresentaram diferença significativa. Assim se o número de reamostragens é menor que o tamanho da amostra, seria recomendado o cálculo do percentil simples devido à sua simplicidade e, nos demais casos, o jackknife. Houve coerência com o teste de razão de verossimilhança e ainda pôde-se concluir que a variação no número de reamostragens não apresentou diferença significativa nos intervalos.

¹ **Comitê Orientador:** Dr. Eduardo Bearzoti – UFLA.(Orientador), Dr. Daniel Furtado Ferreira – UFLA

ABSTRACT

MOURA, Gisele Martins. Use of the resampling techniques jackknife and bootstrap to obtain interval estimates of genetic parameters. In:____. **Type I Error, Power Test and Confidence Intervals With Resampling Techniques in Genetic Models Using Mixtures of Normal Densities.** Lavras: UFLA, 2005. 68p. (Dissertation- Master in Agronomy/ Major in Statistics and Agricultural Experimentation).¹

In plant breeding programs, interval estimates of genetic parameters are as important as point estimates, because they quantify the precision of the estimates and even provide significance test criteria to judge relevant genetic hypotheses. This work aimed at the construction of confidence intervals of parameters associated to polygenes and major genes, using the resampling techniques of jackknife and bootstrap. The bootstrap was implemented using the methods of percentiles and BCa (which has a bias correction and an acceleration parameter). It could be seen that all resampling confidence intervals were adequate and very similar. Therefore, if the number of resamplings is smaller than sample size, then bootstrap-percentile approach could be preferred due to its simplicity; otherwise the jackknife would be more indicated, because it would be faster. The resampling confidence intervals also agreed with the conclusions of the likelihood ratio test with regard the choice of the model.

¹ **Guidance Committee:** Dr. Eduardo Bearzoti – UFLA (Adviser), Dr. Daniel Furtado Ferreira – UFLA

INTRODUÇÃO

A literatura em geral destaca a importância de se conhecer, além da estimativa pontual, a estimativa intervalar de um parâmetro. Triola (1999) afirma que uma simples estimação pontual pode fornecer uma informação limitada a respeito de um parâmetro de interesse, pois não expressa a confiança que se tem acerca da maior ou menor diferença entre a estimativa e o valor real. Bussab e Morettin (1987) destacam, ainda, que a magnitude do erro que se comete com a estimativa pontual pode ser julgada por meio de construção de intervalos de confiança com probabilidades conhecidas de que o intervalo obtido contém o verdadeiro valor do parâmetro.

Neste capítulo procurou-se determinar intervalos de confiança de alguns parâmetros genéticos relativos a poligenes e gene principal, definidos na seção 1.3 do primeiro capítulo, assim como em Silva (2003), que apresenta intervalos de confiança, considerando aproximações normais e erros padrão estimados a partir da matriz de derivadas segundas da função suporte. No entanto, não se conhece a qualidade da aproximação destes erros, nem da distribuição normal dos estimadores. Assim, o ideal seria aplicar técnicas de reamostragem para a obtenção de tais intervalos de confiança.

Diante disso, o objetivo deste trabalho foi construir tais intervalos utilizando as técnicas de reamostragem jackknife e bootstrap, tendo, nesse último, sido utilizados o intervalo percentil simples e BCa. Objetivou-se ainda, implementar a obtenção de tais intervalos no software “Monogen”, desenvolvido por Silva (2003), numa tentativa de sanar esta deficiência no software.

MATERIAL E MÉTODOS

As estimativas dos parâmetros das amostras geradas com os processos de reamostragem jackknife e bootstrap foram obtidas como em Silva (2003), já descrito na seção 1.3, isto é, um processo iterativo que requer um valor inicial, para cada parâmetro. Aqui foram tomadas as estimativas pontuais obtidas em Silva (2003). Assim, como partiu-se de bons “chutes” iniciais, acredita-se não ser necessário um grande número máximo de iterações, portanto adotaram-se apenas 500 iterações.

Vale a pena ressaltar que, primeiro, se deve-se determinar qual modelo se ajusta aos dados; com isso, tem-se de quais parâmetros para devem ser calculadas as respectivas estimativas.

A estimativa intervalar jackknife foi implementada obedecendo à definição apresentada na seção 3.1 do primeiro capítulo (Efron e Tibshirani, 1993), ou seja, era retirada uma observação de cada vez e calculadas as respectivas estimativas jackknife. Em seguida, calculava-se a média dessas estimativas e, assumindo distribuição normal, se constroem os intervalos.

O método bootstrap foi utilizado na versão não-paramétrica, ou seja, assumindo a distribuição uniforme na geração das amostras bootstrap. Os intervalos percentil simples e BCa foram estimados via distribuição bootstrap.

Os procedimentos para a obtenção dos intervalos de confiança bootstrap também seguiram as formas apresentadas na seção 3.2.1 do primeiro capítulo (Efron e Tibshirani, 1993).

Eram obtidas primeiras as reamostragens bootstrap, ressaltando que, no processo de reamostragem, foram preservados os números de indivíduos por geração e também a estrutura das gerações, ou seja, as reamostragens eram feitas dentro de cada geração, a fim de melhor preservar o erro padrão das estimativas dos parâmetros, apesar de essa questão de qual nível hierárquico do modelo

considerado deve ser reamostrado ainda ser não resolvida na literatura (Carlini-Garcia, 2001).

Neste caso, consideraram-se os possíveis valores para o número de reamostragens, 100, 200, 500, 750, 1000, 2000 e 5000, na tentativa de obter uma sugestão sobre qual seria um número suficiente de reamostragens, já que essa também é uma questão bastante discutida (Manly, 1997)

No caso do intervalo percentil simples, eram calculadas as estimativas bootstrap em cada reamostragem, as quais eram ordenadas de forma automática e daí determinado o intervalo contendo 95% dessas estimativas.

Para se obter o intervalo BCa, já com as reamostras bootstrap e jackknife, eram calculadas as variáveis descritas na seção 3.2.2 do primeiro capítulo e determinado o intervalo.

As devidas modificações no software “Monogen” foram implementadas, sendo escritas em linguagem Kylix (Delphi para Linux), agora dando possibilidade ao usuário de realizar, além de testes de hipótese e obter as estimativas pontuais como antes, obter também esses três tipos de intervalos de confiança.

Para ilustrar a construção dos intervalos, utilizou-se o mesmo conjunto de dados de Silva (2003), de um estudo de herança do fenômeno da partenocarpia em abobrinha (*Cucurbita pepo*). O experimento para tal foi realizado no setor de olericultura da Universidade Federal de Lavras, MG, considerando as seis populações já mencionadas P_1 , P_2 , F_1 , F_2 , RC_1 e RC_2 , em que os genitores contrastantes considerados foram a variedade *Caserta* (P_1) e a variedade *Whitaker* (P_2). As tabelas com o número das plantas em cada geração e o valor das notas atribuídas a ocorrência de partenocarpia de cada planta conforme as regras usadas no estudo, podem ser encontradas em Silva (2003).

RESUTADOS E DISCUSSÃO

Silva (2003) concluiu que o modelo que melhor se ajusta aos dados é o modelo 7, ou seja, o que contém os parâmetros: constante de referência, variância ambiental, componente aditivo do gene principal e componente de dominância do gene principal. Assim, aqui também serão considerados tais parâmetros, já que se trata do mesmo conjunto de dados.

Na Tabela 1 encontram-se o valor das estimativas pontuais obtidas por Silva (2003) e os resultados do método de reamostragem jackknife.

TABELA 1 Resultados do método de reamostragem jackknife

Parâmetros	Estimativa pontual	Estimativa jackknife	Erro padrão jackknife	Intervalo de confiança jackknife
μ	2,854684	2,848451	0,049721	[2,757240; 2,952150]
σ^2	0,805040	0,815855	0,096874	[0,615146; 0,994895]
A	1,276216	1,272759	0,067322	[1,144270; 1,408174]
D	0,362283	0,369579	0,135795	[0,096110; 0,628429]

Pode-se observar, na Tabela 1, que o jackknife sugere haver viés, porém, pequeno, já que se têm estimativas jackknife muito próximas das obtidas com os dados reais.

Os intervalos de confiança bootstrap, percentil simples estão apresentados nas Tabelas 2 e 3 e os intervalos BCa estão nas Tabelas 4 e 5, que os apresentam para as devidas variações do número de reamostragens.

TABELA 2 Intervalos de confiança percentil simples para os parâmetros μ e σ^2

Número de reamostragens	PARÂMETROS	
	μ	σ^2
100	[2,740203 ; 2,937702]	[0,619136 ; 0,950535]
200	[2,751883 ; 2,941480]	[0,622933 ; 0,950535]
500	[2,750676 ; 2,953924]	[0,619136 ; 0,979052]
750	[2,750674 ; 2,954929]	[0,622064 ; 0,976905]
1000	[2,750674 ; 2,947684]	[0,626489 ; 0,979052]
1500	[2,750486 ; 2,946359]	[0,627312 ; 0,976698]
2000	[2,750225 ; 2,943771]	[0,628342 ; 0,973262]
5000	[2,753065 ; 2,945696]	[0,626489 ; 0,982970]

TABELA 3 Intervalos de confiança percentil simples para os parâmetros A e D.

Número de reamostragens	PARÂMETROS	
	A	D
100	[1,119361 ; 1,371579]	[0,069436 ; 0,657991]
200	[1,151878 ; 1,378704]	[0,069436 ; 0,616384]
500	[1,150259 ; 1,401380]	[0,084169 ; 0,647567]
750	[1,145240 ; 1,400459]	[0,092795 ; 0,629766]
1000	[1,134120 ; 1,399590]	[0,091498 ; 0,639393]
1500	[1,139646 ; 1,398076]	[0,091498 ; 0,634705]
2000	[1,141294 ; 1,398136]	[0,094760 ; 0,642852]
5000	[1,141089 ; 1,399530]	[0,100563 ; 0,633877]

TABELA 4 Intervalos de confiança BCa para os parâmetros μ e σ^2

Número de reamostragens	PARÂMETROS	
	μ	σ^2
100	[2,759796 ; 2,937702]	[0,622933 ; 0,950535]
200	[2,763428 ; 2,943133]	[0,644763 ; 0,966829]
500	[2,751883 ; 2,953924]	[0,627059 ; 0,989916]
750	[2,750676 ; 2,955524]	[0,634662 ; 0,984892]
1000	[2,750676 ; 2,947735]	[0,640067 ; 0,989916]
1500	[2,752774 ; 2,949169]	[0,639755 ; 0,984892]
2000	[2,750674 ; 2,944788]	[0,638815 ; 0,976905]
5000	[2,750676 ; 2,943786]	[0,632362 ; 0,988306]

TABELA 5 Intervalos de confiança BCa para os parâmetros A e D.

Número de reamostragens	PARÂMETROS	
	A	D
100	[1,151878 ; 1,371579]	[0,032110 ; 0,554289]
200	[1,154532 ; 1,378704]	[0,032110 ; 0,597638]
500	[1,152282 ; 1,402150]	[0,066845 ; 0,616384]
750	[1,150259 ; 1,402150]	[0,084169 ; 0,629766]
1000	[1,138114 ; 1,400459]	[0,090673 ; 0,634705]
1500	[1,142114 ; 1,399590]	[0,089391 ; 0,633772]
2000	[1,145291 ; 1,400798]	[0,096811 ; 0,645365]
5000	[1,146362 ; 1,401568]	[0,107931 ; 0,642852]

Pode-se destacar que os intervalos BCa concordam que o viés deve ser pequeno, já que ele possui um parâmetro que corrige o possível viés do percentil simples; porém, eles, em geral, apresentaram pouca diferença.



Pode-se destacar que houve coerência com o teste de razão de verossimilhança, que diz que o modelo 7 é o modelo correto, já que nenhum desses intervalos contém o valor zero; assim, são estatisticamente não nulos.

Os três tipos de intervalos não apresentaram diferença expressiva. Assim, se o número de reamostragens for um número menor que o tamanho da amostra, seria recomendado o uso do intervalo percentil simples, já que é o de mais fácil obtenção, como destaca Perreira (2000). Porém, caso contrário, a melhor opção seria o jackknife.

Neste exemplo, que é bastante característico, pôde-se observar que o número de reamostragens não apresentou influência significativa.

O ideal seria fazer estudos de simulações para confirmar a eficácia de tais intervalos. Já os resultados verificados aqui dizem respeito a este exemplo, que é um caso muito particular.

CONCLUSÕES

Os intervalos jackknife e BCa sugerem que houve viés, porém, pequeno. Houve coerência com o teste de razão de verossimilhança. O intervalo de confiança jackknife e os intervalos bootstrap percentil simples e BCa apresentaram-se como uma alternativa satisfatória na estimação intervalar desses parâmetros genéticos, no exemplo em questão.

CONCLUSÕES GERAIS

O teste de razão de verossimilhança para testar a existência de poligenes e/ou gene principal mostrou-se uma excelente alternativa, já que apresentou taxas de erro tipo I abaixo do valor nominal, associado a um valor de poder muito alto.

No exemplo em questão, os intervalos de confiança baseados em técnicas de reamostragem jackknife, e os intervalos bootstrap percentil simples e BCa se mostraram uma boa alternativa na estimação intervalar dos parâmetros genéticos.

REFERÊNCIAS BIBLIOGRÁFICAS

- BUSSAB, W. O.; MORRETTIN, P. A. **Estatística Básica**. 4. ed. São Paulo: Atual, 1987. 321 p.
- CARLINI-GARCIA, L. A. **Uso do método de reamostragem bootstrap na estimação de parâmetros genéticos populacionais**. 2001. 162 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba.
- EFRON, B.; TIBSHIRANI, R. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993. 436 p.
- PEREIRA, J. E.; SILVA, J. F. V.; DIAS, W. F.; SOUZA, G. S. Intervalo de confiança “bootstrap” como ferramenta para classificar raças do nematóide de cisto da soja. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 2, p. 271-275, fev. 2000.
- MANLY, B. F. J. **Randomization, bootstrap and Monte Carlo methods in biology**. 2. ed. London: Chapman-Hall, 1998. 399 p.
- SILVA, W. P. **Estimadores de máxima verossimilhança em misturas de densidades normais: uma aplicação em genética**. 2003. 60 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.
- TRIOLA, M. F. **Introdução a estatística**. 7. ed. Rio de Janeiro: LTC, 1999. 410 p.