

COMPARAÇÃO ENTRE MÉTODOS DE ESTIMAÇÃO DO COEFICIENTE DE ENDOGÂMIA COM DADOS DE FREQUÊNCIAS ALÉLICAS EM UMA POPULAÇÃO DIPLÓIDE¹

Comparison among inbreeding coefficient estimation methods in a diploid population with two alleles

Joel Augusto Muniz², Gabriela Quandt Costa³, Ricardo Luis Reis⁴, Ruben Delly Veiga⁵

RESUMO

Neste trabalho, objetivou-se comparar três estimadores do coeficiente de endogamia, F , em uma população diplóide com dois alelos, utilizando-se dados de frequências alélicas em amostras de indivíduos, com diferentes tamanhos obtidas em populações simuladas, por meio do software SAS. Foi avaliado o estimador de F , obtido pela análise de variância de frequências alélicas, o estimador considerando o método dos momentos e o estimador pelo método da máxima verossimilhança. Os resultados encontrados para a média e variância os estimadores, a partir de 1000 estimativas de F , calculadas para cada tamanho de amostra, mostraram que os três estimadores são tendenciosos. Entretanto, de maneira geral, observou-se que o estimador considerando a análise de variância foi menos tendencioso e apresentou menor variância, quando o coeficiente de endogamia da população foi alto. Para tamanho de amostra superior a 50, os três estimadores tiveram comportamento semelhante, independente da frequência alélica e da endogamia da população.

Termos para indexação: Método dos momentos, método da máxima verossimilhança, análise de variância de frequências alélicas, tamanho amostral, viés, coeficiente de endogamia.

ABSTRACT

The present work evaluated the properties of three estimators of the inbreeding coefficient, F , in a diploid population with two alleles, using data of gene frequencies in individuals from random samples obtained from populations simulated through the SAS. We evaluated the estimators of F obtained by variance analysis of allelic frequencies, obtained by moment method, and estimator obtained by maximum likelihood method. The analysis of the means and variances of the estimators, obtained from 1000 estimates of F , calculated for each sample size, demonstrated that the three estimators were biased. However, it was observed that the estimator obtained from univariate analysis was less biased and presented smaller variance, when the inbreeding coefficient in the population was elevated, while for populations with low inbreeding, the variance of the estimator obtained by the multivariate analysis was smaller.

Index terms: moment method, maximum likelihood method, variance analysis of allelic frequencies, sample size, bias, inbreeding coefficient.

(Recebido em 22 de junho de 2007 e aprovado em 9 de janeiro de 2009)

INTRODUÇÃO

A estimação do coeficiente de endogamia de uma população diplóide utilizando dados de frequências alélicas, de acordo com Vencovsky (1992) e Weir (1996) considera a variável indicadora y que vale 1 quando determinado alelo, por exemplo A_1 de um loco, está no indivíduo e vale 0 quando este alelo está ausente e presentes os alelos A_2, A_3, \dots, A_n . Para as frequências dessa variável pode-se estimar os parâmetros genéticos de população, utilizando métodos clássicos de estimação. No processo de estimação ocorrem algumas dificuldades, pois, de acordo com a teoria estatística, a definição das

propriedades de estimadores se baseia na distribuição envolvida, e é sabido que a variável y associada às frequências alélicas não apresenta distribuição normal, dificultando a obtenção da distribuição dos estimadores. Além disso, em alguns casos, a estimação envolve quocientes entre variáveis aleatórias que também nem sempre apresentam distribuições conhecidas, exigindo a utilização de estudo de aproximações por meio de série de Taylor.

Segundo Cockerham (1969), endogamia, variâncias de frequências alélicas, tamanho efetivo de população são termos comuns no estudo da genética de populações, tendo os conceitos e a maior parte da teoria sido introduzidos pelos

¹Trabalho desenvolvido com o apoio do CNPq.

²Universidade Federal de Lavras/UFLA – Departamento de Ciências Exatas/DEX – Lavras, MG – joamuniz@ufla.br

³Universidade Federal de Lavras/UFLA – Departamento de Ciências Exatas/DEX – Lavras, MG

⁴Universidade Federal de Lavras/UFLA – Departamento de Ciências Exatas/DEX – Lavras, MG

⁵Universidade Federal de Lavras/UFLA – Departamento de Ciências Exatas/DEX – Lavras, MG

trabalhos clássicos de Wright (1921) e Fisher (1949). De acordo com Falconer (1964), a endogamia ocorre quando se tem o acasalamento entre indivíduos que são relacionados por ascendência, tendo como primeiro efeito uma mudança nas frequências genotípicas de Hardy-Weinberg decorrente de um aumento na frequência de genótipos homozigóticos em detrimento das frequências de genótipos heterozigóticos. Para o autor, o coeficiente de endogamia F mede a probabilidade de dois genes, em qualquer loco de um indivíduo serem originados da cópia de apenas um gene na geração anterior.

Hartl & Clark (1989) afirmam que, em termos biológicos, o coeficiente de endogamia F expressa a redução fracionária da heterozigosidade, em relação a uma população com acasalamento aleatório. No caso de um loco com dois alelos, as frequências genotípicas de AA, Aa e aa podem ser definidas em relação ao coeficiente de endogamia pela função $(1-F)(p^2, 2pq, q^2) + F(p, 0, q)$ sendo p , a frequência do alelo A e $q = 1-p$, a frequência do alelo a. Para os autores, esses resultados facilitam o entendimento e a comparação das frequências genotípicas no princípio de Hardy-Weinberg. Quando $F = 0$, tem-se acasalamento aleatório, ou seja, não existe endogamia e as frequências genotípicas da população estarão de acordo com o equilíbrio de Hardy-Weinberg. Quando F assume qualquer outro valor entre 0 e 1, as frequências se alteram, sendo que para $F = 1$, tem-se endogamia total e a população só apresentará genótipos homozigóticos AA e aa, respectivamente nas frequências p e q .

Wright (1965) discute o relacionamento entre gametas, mostrando o significado do coeficiente de endogamia e de parâmetros relacionados a pares de gametas em geral, realçando o fato de que eles podem ser interpretados tanto como coeficientes de correlação, quanto como a probabilidade dos genes idênticos serem de mesma origem. No caso de populações naturais hierarquicamente subdivididas, três parâmetros de descrição ou índices de fixação foram propostos em termos da população total (T), de subpopulações (S) e de indivíduos (I). Os parâmetros são os seguintes:

- F_{IT} - expressa a correlação entre os gametas que se unem para produzir os indivíduos em relação aos gametas da população total.
- F_{IS} - expressa a média das correlações, sendo cada uma delas proveniente dos gametas que se unem em cada subpopulação em relação aos gametas desta subpopulação.
- F_{ST} - expressa a correlação entre os gametas ao acaso dentro da subpopulação em relação aos gametas da população total.

Segundo o autor, o parâmetro F_{IT} reúne informações dos outros dois, existindo a seguinte relação: $1 - F_{IT} = (1 - F_{IS})(1 - F_{ST})$.

Weir & Cockerham (1984) consideram, para o caso de um dos alelos de um loco, as seguintes definições e notações: F , a correlação entre os genes dentro de indivíduos ou endogamia; q , a correlação entre os genes de diferentes indivíduos da mesma população ou coancestria e f , a correlação entre genes dentro de indivíduos dentro de populações. Os três parâmetros correspondem às estatísticas F de Wright da seguinte forma: $F = F_{IT}$, $q = F_{ST}$ e $f = F_{IS}$ que estão relacionados pela expressão: $f = (F - \theta)/(1 - \theta)$.

Reynolds et al. (1984), estudaram a estimação do coeficiente de coancestria, no caso de uma população com alelos múltiplos, sugerindo um estimador com base na média de todos os u alelos e outro considerando a análise conjunta, envolvendo os u alelos. Os dois estimadores de q propostos foram avaliados por Weir & Cockerham (1984) sendo que o segundo estimador foi considerado melhor que o primeiro, por apresentar resultados mais consistentes.

Vencovsky (1992) descreve a técnica de análise de variância com frequências gênicas, quando os dados estão organizados em progênies e discute sua utilização para descrever a estrutura genética de populações diplóides. O autor considera modelo aleatório, de acordo com Cockerham (1969), onde se admite que a diversidade entre populações ocorre, fundamentalmente, pela variação genética nas gerações passadas. São apresentados os parâmetros genéticos e suas relações com os componentes de variância associados à análise de variância das frequências gênicas. Os parâmetros são estimados por meio dos componentes de variância, utilizando-se a aproximação que considera a estimativa de um quociente como sendo o quociente entre as estimativas. Oliveira et al. (2009) relatam as dificuldades de estudar as propriedades estatísticas de estimador que envolve o quociente de variáveis aleatórias por não apresentar distribuição estatística definida.

Weir (1996) apresenta uma discussão geral sobre os métodos de estimação de parâmetros genéticos com base em dados de frequências alélicas. Entre os diversos métodos o autor destaca o método dos momentos, o método da máxima verossimilhança e a análise de variância das frequências alélicas. No caso da técnica da análise de variância, o autor aborda o caso de organismos haplóides bem como populações diplóides com modelos hierárquicos de até quatro níveis.

Hill et al. (1995) desenvolveram métodos para a estimação do coeficiente de endogamia em populações com um ou mais locus, com vários alelos. Os métodos discutidos foram ilustrados com dados de frequências alélicas associados a ovos do parasita causadores da malária humana (*Plasmodium falciparum*) amostrados em mosquitos

vetores. Ayres & Balding (1998) propuseram um método Bayesiano para testar desvios da proporções de Hardy-Weinberg em modelos genéticos com endogamia. Os autores avaliaram ainda o método de Neir & Chesser (1983), o de Robertson & Hill (1984) bem como o método da máxima verossimilhança, nada sendo discutido sobre as variâncias dos estimadores. Foi feita a ilustração por meio de simulação comparando a estimação num modelo com coeficientes de endogamia igual a 5%, considerando-se amostras de tamanhos 200 e 1000 e populações com dois, seis e quinze alelos, apresentando frequências diferentes, obtendo-se resultados mais consistentes com o aumento do tamanho da amostra.

Rousset & Raymond (1995) propuseram um teste que considera as estimativas de máxima verossimilhança do coeficiente de endogamia das frequências alélicas e avaliaram suas propriedades quanto ao poder e consistência. Segundo os autores, o teste proposto é mais apropriado que o teste de χ^2 para testar os desvios das proporções de Hardy-Weinberg quando a hipótese alternativa é a deficiência de heterozigotos.

Muniz et al. (1996) estudaram as propriedades dos estimadores do coeficiente de endogamia e da taxa de fecundação cruzada obtidos pela análise de variância com dados de frequências alélicas em populações diplóides. Os autores observaram que os estimadores são tendenciosos.

Muniz et al. (1997a) compararam três métodos de estimação da variância do estimador do coeficiente de endogamia obtido na análise de variância das frequências alélicas em uma população diplóide. Resultados de simulação mostraram que os métodos propostos apresentam valores semelhantes e satisfatórios quando a frequência alélica da população estiver entre 0,3 e 0,7 o coeficiente de endogamia da população for inferior a 0,5 e se trabalhar com pelo menos 30 indivíduos. Muniz et al. (1997b) estudaram a estimação do coeficiente de endogamia pelo método dos momentos em populações diplóides com alelos múltiplos. Os autores verificaram que o estimador com base na análise de variância conjunta envolvendo todos os alelos foi menos tendencioso que o estimador usando a média das análises de cada alelo.

Muniz et al. (1999), estudando a estimação do coeficiente de endogamia em uma população diplóide, avaliaram a distribuição do quociente dos quadrados médios entre indivíduos e entre genes dentro de indivíduos, verificando que o teste de F da análise de variância pode ser utilizado para testar a nulidade do coeficiente de endogamia quando a frequência alélica estiver entre 0,3 e 0,7 trabalhando-se com 30 indivíduos e, entre 0,20 e 0,80, com 100 indivíduos. Os resultados teóricos encontrados foram ilustrados por meio de valores numéricos obtidos em estudo de simulação.

Muniz et al. (2001a) estudaram a distribuição do coeficiente entre quadrados médios na análise de variância de frequências alélicas de indivíduos em populações haplóides, avaliando o teste F para testar a nulidade do coeficiente de coancestria. Os autores observaram que o teste F pode ser utilizado quando se trabalha com pelo menos cinco populações com frequência alélica média entre 0,3 e 0,7 utilizando-se no mínimo 50 indivíduos. Muniz et al. (2001b) avaliaram as propriedades do estimador do coeficiente de coancestria em populações haplóides com dois alelos. Os autores desenvolveram expressões por série de Taylor e avaliaram as propriedades assintóticas. Resultados numéricos de simulação validaram as expressões algébricas propostas.

Muniz et al. (2008) compararam métodos de estimação do coeficiente de endogamia em populações diplóides com três alelos. Os autores observaram que todos os métodos foram tendenciosos, mas a estimação por meio da análise multivariada, envolvendo os alelos, apresenta menor variância em populações com endogamia baixa do que a estimação pela análise conjunta dos alelos e pela média dos alelos. Reis et al. (2009) afirmaram que uma alternativa para estudos de simulação de dados muito utilizada em trabalhos de melhoramento genético é a técnica de Monte Carlo, que gera números aleatórios segundo uma distribuição de probabilidade. Silva et al. (2008) utilizaram estudo de simulação Monte Carlo para avaliação de testes sobre a igualdade de matrizes de covariâncias de populações.

Mood et al. (1974) discutem os métodos clássicos de estimação, ou seja, o método dos momentos, o método dos mínimos quadrados e o método da máxima verossimilhança. Os autores destacam que o método da máxima verossimilhança, apesar de produzir, em alguns casos, estimadores tendenciosos, apresenta propriedades ótimas quando de trabalha com grandes amostras, uma vez que os estimadores obtidos são funções estatísticas suficientes. Assim, assintoticamente, o vetor de estimativas de máxima verossimilhança apresenta distribuição normal multivariada com vetor de médias e matrizes de variância e covariância \mathbf{w}^{-1} , sendo \mathbf{w} a matriz de informação de Fisher, obtida pelas derivadas de segunda ordem do logaritmo da função de verossimilhança envolvida.

Em estudos de genética de populações é importante conhecer as propriedades dos estimadores do coeficiente de endogamia para que se possa adotar o método de estimação mais adequado. Assim, o presente trabalho visou a comparar três métodos de estimação do coeficiente de endogamia: a estimação por meio da análise de variância das frequências alélicas, o método dos momentos e o método da máxima verossimilhança procurando-se avaliar o comportamento dos estimadores em amostras com diferentes número de indivíduos, extraídas de populações diplóides com diversos graus de endogamia.

MATERIAL E MÉTODOS

No caso de uma amostra de indivíduos de uma população diplóide com dois alelos, a variável indicadora da ocorrência de determinado alelo é descrita pelo modelo aleatório de Cockerham (1969): $y_{ij} = p + a_i + g_{(j)i}$. Nesse modelo: y_{ij} é a frequência do alelo j dentro do indivíduo i ; p é a frequência paramétrica do alelo A na população; a_i é o efeito do indivíduo i , com $i = 1, 2, \dots, n$ e $g_{(j)i}$ é o efeito do alelo j dentro do indivíduo i , com $j = 1, 2$.

Admitindo a metodologia descrita por Vencovsky (1992), obtém-se a análise de variância para as frequências alélicas de uma amostra de n indivíduos:

Causa de Variação	GL	QM	E(QM)
Indivíduos	$n-1$	QMI	$p(1-p)[(1-F) + 2F]$
Alelos dentro de indivíduos	n	QMA	$p(1-p)[(1-F)]$

A estimação de F com base nas esperanças de quadrados médios da análise de variância é dada pela

expressão: $\hat{F}_1 = \frac{QMI - QMA}{QMI + QMA}$. Esse estimador, conforme resultados obtidos por Muniz et al. (1997a) é tendencioso e apresenta variância dada pela expressão:

$\text{Var}(\hat{F}_1) = \frac{(1-F)^2}{n}$, sendo n o número de indivíduos e F o coeficiente de endogamia da população.

Weir (1996) descreve a estimação do coeficiente de endogamia pelo método dos momentos numa população com u alelos. Para uma população com apenas dois alelos o estimador proposto apresenta a seguinte expressão:

$$\hat{F}_2 = 1 - \frac{(n-1) \frac{n_{Aa}}{n}}{2n \hat{p}_A (1 - \hat{p}_A) - \frac{n_{Aa}}{n}}$$

sendo: n_{Aa} o número de indivíduos heterozigóticos na amostra, \hat{p}_A a estimativa da frequência do alelo A na amostra e n número de indivíduos da amostra. O autor desenvolve ainda estudo de série de Taylor, obtendo aproximação para a variância do estimador para populações com u alelos. No caso de populações com apenas dois alelos a variância do estimador tem expressão teórica dada

por: $\text{Var}(\hat{F}_2) = \frac{2p_A(1-p_A)}{n}$, sendo p_A a frequência do alelo A na população.

Weir (1996) desenvolve um estudo com base no método de Bailey (1951), para a estimação de parâmetros genéticos em dados com estrutura multinomial e propõe um estimador para o coeficiente de endogamia de uma população com dois alelos, pelo método da máxima verossimilhança definido pela expressão:

$$\hat{F}_3 = 1 - \frac{n_{Aa}}{2n \hat{p}_A (1 - \hat{p}_A)}$$

sendo n_{Aa} o número de indivíduos heterozigóticos na amostra, \hat{p}_A a estimativa da frequência do alelo A na amostra e n o número de indivíduos da amostra. O autor considerou os valores esperados das contagens genotípicas e obteve a variância do estimador por:

$$\text{Var}(\hat{F}_3) = \frac{1}{n}(1-F)^2(1-2F) + \frac{F(1-F)(2-F)}{2np_A(1-p_A)}$$

sendo F o coeficiente de endogamia da população, n o número de indivíduos e p_A a frequência populacional do alelo A .

Para obtenção da distribuição dos estimadores, cálculo da média e variância e avaliação das propriedades de não-tendenciosidade e de variância mínima, foram feitas simulações de populações endogâmicas de 1000 indivíduos com dois alelos, endogamia e frequências alélicas, variando de 0,1 no intervalo entre 0 e 1, utilizando-se programas desenvolvidos no Software Statistical Analysis System (SAS). Foram obtidas 121 populações, nas quais foram simulados 1000 experimentos para cada um dos tamanhos de amostras considerados ($n = 10, 20, 30, 40, 50, 100$ e 200 indivíduos), totalizando, assim, 847.000 experimentos. Os experimentos foram simulados utilizando-se a função RANUNI (SAS Institute, 1990) que gera valores para uma variável aleatória uniformemente distribuída entre 0 e 1, obtendo, assim, amostras dentro de cada população. Em cada experimento foram calculadas as estimativas do coeficiente de endogamia F , utilizando-se as expressões \hat{F}_1 , \hat{F}_2 e \hat{F}_3 dos estimadores apresentados, considerando-se as frequências alélicas dos n indivíduos amostrados. Foram calculados também os valores paramétricos para as variâncias dos três estimadores com base nas expressões apresentadas na literatura para comparação com os valores experimentais.

RESULTADOS E DISCUSSÃO

Os valores de média e variância dos três estimadores foram obtidos, nos $N = 1000$ experimentos simulados, em

cada situação, envolvendo tamanho de amostra, valor do coeficiente de endogamia da população e frequência alélica populacional, utilizando-se o procedimento univariado (Proc. Univariate) do SAS Institute (1990). Os dados foram organizados em tabelas, de forma a facilitar e orientar a visualização dos resultados experimentais simulados. Não foram apresentados todos os resultados obtidos em razão do grande volume de informação, optando-se pela apresentação de situações representativas, onde p e F assumem valores baixos, intermediários e altos, correspondendo aos casos específicos de frequência alélica $p = 0,1; 0,5$ e $0,9$ e endogamia populacional $F = 0,1; 0,5$ e $0,9$ e amostras de tamanho $n = 10, 30, 50, 100$ e 200 .

Os resultados apresentados na Tabela 1 indicam que os três estimadores foram tendenciosos, sendo que F_1 (análise de variância) foi o menos tendencioso, especialmente para populações com frequência alélica baixa onde se percebe que amostras de no mínimo 50 indivíduos fornecem resultados satisfatórios. O estimador mais tendencioso foi o F_3 (método da máxima verossimilhança), o que pode ser observado em todas populações avaliadas. Os valores médios da simulação com base em 1000 experimentos comprovam que o método dos momentos (F_2) só foi eficiente, em todas populações estudadas, com o uso de amostras de 200 indivíduos, sendo que para amostras com menor número de indivíduos ele se mostrou o mais tendencioso, conforme ilustrado nas Figuras (1, 2 e 3). Esses resultados concordam com a afirmação de Mood et al. (1974) de que o método de máxima verossimilhança, em geral, só apresenta boas propriedades quando se utilizam grandes amostras. Para populações com frequência alélica 0,5 o estimador F_1 apresentou médias de estimativas próximas do valor paramétrico com amostras a partir de 30 indivíduos. Os resultados encontrados concordaram com aqueles obtidos por Muniz et al. (1997a), destacando-se o fato de que as

variâncias dos três estimadores mostraram-se semelhantes quando foram utilizadas amostras com mais de 30 indivíduos. De modo geral, observou-se que todos os três estimadores apresentaram viés negativo, uma vez que os valores das médias dos mesmos ficaram sempre abaixo do valor teórico do coeficiente de endogamia em todas as situações simuladas. Isso se deve ao fato de que as expressões dos estimadores envolvem quocientes entre variáveis aleatórias. Esses quocientes não possuem distribuição conhecida, sendo avaliados por aproximação de série de Taylor, onde o resto do desenvolvimento da série define o sinal do viés. Com o aumento do tamanho da amostra, houve redução do viés, conforme verificado nas Figuras (1, 2 e 3). Esta observação concorda com os resultados obtidos por Muniz et al. (1997b) e Muniz (2008).

Os resultados da simulação indicaram que o estimador F_1 apresentou, em todas populações consideradas, valores de variância acima dos valores paramétricos calculados com base na expressão sugerida por Muniz et al. (1997a), conforme valores da Tabela 2. Observou-se ainda que, com o aumento da endogamia da população, as diferenças observadas se tornam maiores. Para populações com frequência alélica 0,5 e endogamia 0,1 porém, a variância de F_1 forneceu valores próximos dos valores esperados.

A expressão proposta por Weir (1996) produz valores iguais em populações com frequência alélica 0,1 e 0,9. A variância do estimador F_2 apresentou valores relativamente próximos dos valores populacionais esperados (Tabela 3) apenas nas populações com endogamia alta, indicando que a expressão sugerida pelo autor só é adequada para estas situações. Quanto ao estimador F_3 , sua variância apresentou valores próximos dos valores populacionais esperados conforme os resultados da Tabela 4, com amostras a partir de 30 indivíduos, comprovando-se a validade da expressão sugerida pelo autor.

Tabela 1 – Médias e variâncias dos estimadores do coeficiente de endogamia obtidas por três métodos (F_1 – método da análise de variância; F_2 – método dos momentos; F_3 – método da máxima verossimilhança) calculadas com base em 1000 experimentos simulados, utilizando-se amostras (n) de tamanhos diferentes, extraídas de populações com dois alelos, com diversas combinações de freqüências gênicas e coeficientes de endogamia.

p	F	n	μ (F_1)	Var (F_1)	μ (F_2)	Var (F_2)	μ (F_3)	Var (F_3)	
0,1	0,1	10	0,0708	0,0816	0,0118	0,0926	0,0195	0,0849	
		30	0,0717	0,0434	0,0552	0,0454	0,0549	0,0435	
		50	0,0930	0,0303	0,0843	0,0314	0,0833	0,0305	
		100	0,0967	0,0162	0,0925	0,0165	0,0918	0,0162	
		200	0,0978	0,0077	0,0957	0,0078	0,0953	0,0077	
	0,5	10	0,3823	0,1798	0,3476	0,2039	0,3449	0,1966	
		30	0,4415	0,0887	0,4352	0,0922	0,4300	0,0909	
		50	0,4872	0,0494	0,4840	0,0505	0,4800	0,0503	
		100	0,4849	0,0212	0,4834	0,0214	0,4812	0,0214	
		200	0,4957	0,0107	0,4950	0,0108	0,4938	0,0108	
	0,9	10	0,8536	0,1013	0,8279	0,1228	0,8257	0,1225	
		30	0,8688	0,0426	0,8691	0,0436	0,8667	0,0442	
		50	0,8867	0,0158	0,8864	0,0160	0,8847	0,0163	
		100	0,8915	0,0063	0,8915	0,0063	0,8905	0,0064	
		200	0,8930	0,0029	0,8929	0,0029	0,8925	0,0029	
	0,5	0,1	10	0,0773	0,1035	0,0207	0,1289	0,0303	0,1043
			30	0,0913	0,0346	0,0762	0,0369	0,0751	0,0348
			50	0,0925	0,0189	0,0838	0,0196	0,0827	0,0189
			100	0,0969	0,0099	0,0927	0,0100	0,0920	0,0099
			200	0,0985	0,0048	0,0964	0,0048	0,0960	0,0048
0,5		10	0,4697	0,0809	0,4486	0,0924	0,4322	0,0881	
		30	0,4894	0,0259	0,4845	0,0269	0,4769	0,0267	
		50	0,4930	0,0142	0,4902	0,0145	0,4854	0,0145	
		100	0,4980	0,0078	0,4967	0,0079	0,4942	0,0079	
		200	0,4982	0,0038	0,4976	0,0039	0,4963	0,0039	
0,9		10	0,8884	0,0215	0,8861	0,0225	0,8777	0,0254	
		30	0,8971	0,0067	0,8968	0,0067	0,8939	0,0071	
		50	0,8950	0,0041	0,8949	0,0042	0,8930	0,0043	
		100	0,8972	0,0020	0,8972	0,0020	0,8962	0,0020	
		200	0,8982	0,0009	0,8981	0,0009	0,8977	0,0009	
0,9		0,1	10	0,0518	0,0816	0,0281	0,0979	0,0075	0,0913
			30	0,0694	0,0439	0,0530	0,0459	0,0528	0,0440
			50	0,0876	0,0307	0,0787	0,0318	0,0779	0,0308
			100	0,0914	0,0142	0,0871	0,0144	0,0865	0,0142
			200	0,0924	0,0075	0,0904	0,0076	0,0900	0,0075
	0,5	10	0,3916	0,1945	0,3591	0,2163	0,3573	0,2091	
		30	0,4444	0,0935	0,4383	0,0973	0,4332	0,0959	
		50	0,4621	0,0553	0,4586	0,0566	0,4547	0,0562	
		100	0,4860	0,0211	0,4846	0,0214	0,4823	0,0213	
		200	0,4833	0,0113	0,4826	0,0113	0,4814	0,0113	
	0,9	10	0,8473	0,0986	0,8361	0,1145	0,8336	0,1143	
		30	0,8781	0,0384	0,8769	0,0403	0,8746	0,0409	
		50	0,8696	0,0267	0,8693	0,0272	0,8676	0,0275	
		100	0,8850	0,0072	0,8849	0,0072	0,8840	0,0073	
		200	0,8900	0,0032	0,8900	0,0032	0,8895	0,0032	

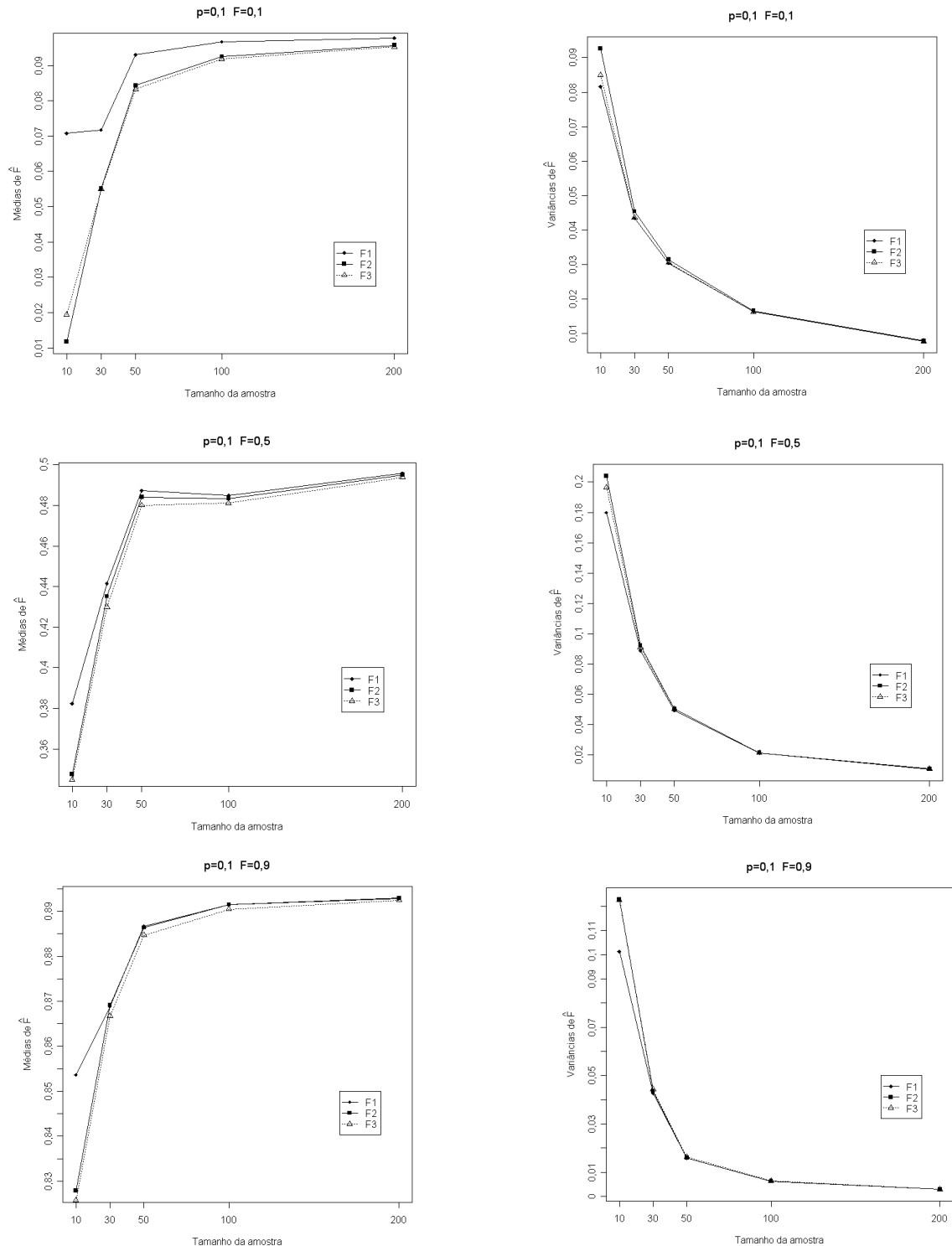


Figura 1 – Médias e variâncias dos estimadores do coeficiente de endogamia obtidas por três métodos (F_1 – método da análise de variância; F_2 – método dos momentos; F_3 – método da máxima verossimilhança), em populações com frequência alélica 0,1 e diferentes coeficientes de endogamia.

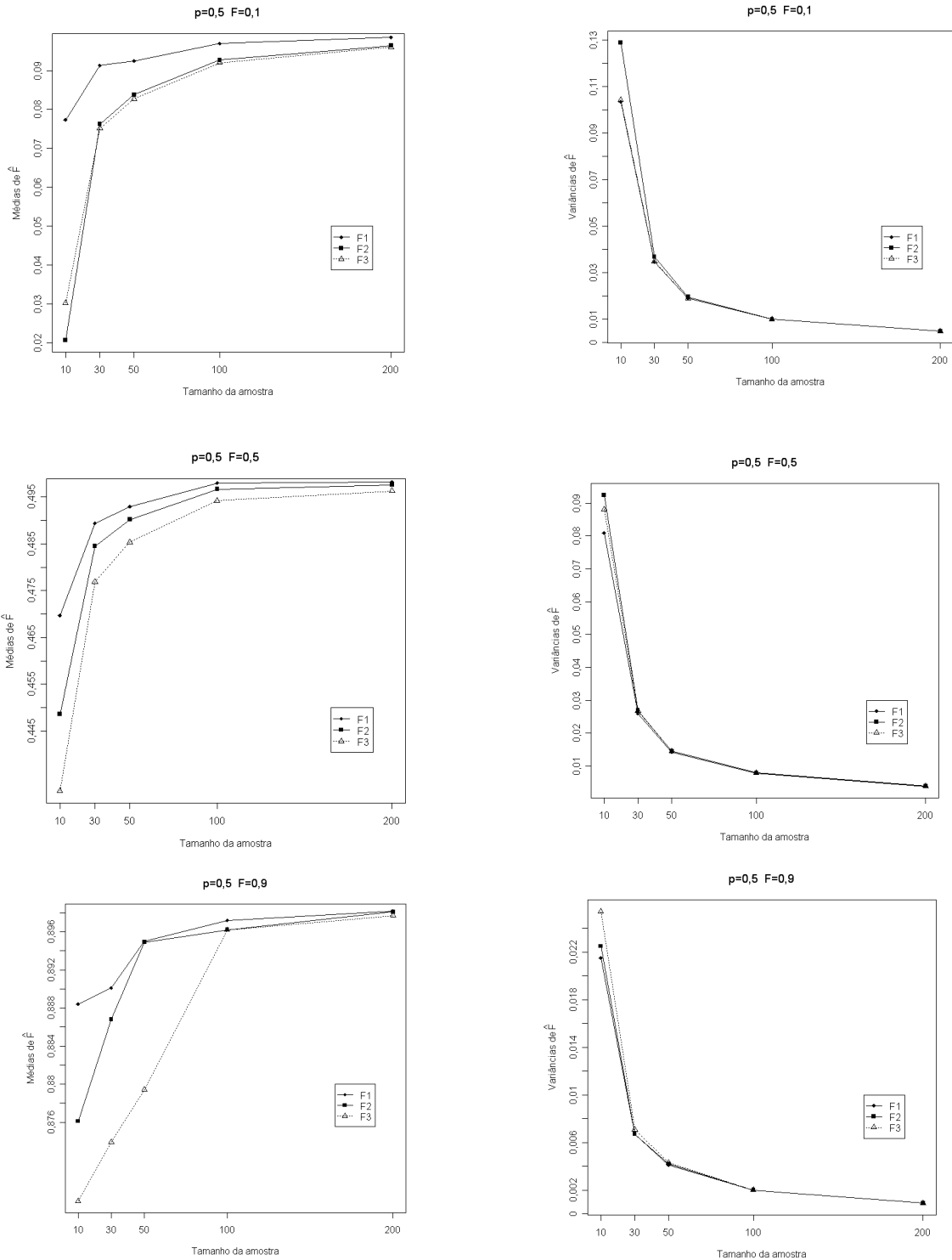


Figura 2 – Médias e variâncias dos estimadores do coeficiente de endogamia obtidas por três métodos (F_1 – método da análise de variância; F_2 – método dos momentos; F_3 – método da máxima verossimilhança), em populações com frequência alélica 0,5 e diferentes coeficientes de endogamia.

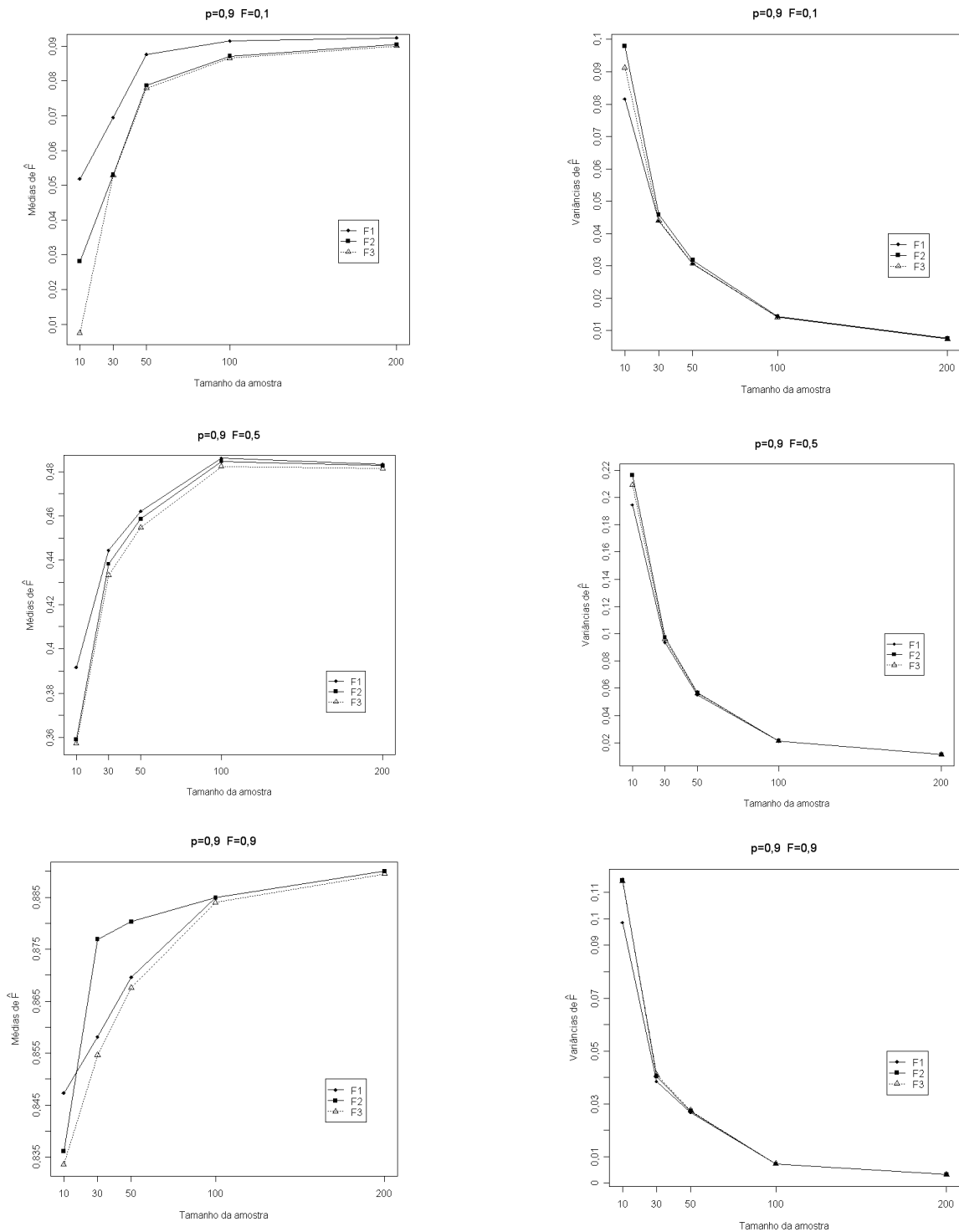


Figura 3 – Médias e variâncias dos estimadores do coeficiente de endogamia obtidas por três métodos (F_1 – método da análise de variância; F_2 – método dos momentos; F_3 – método da máxima verossimilhança), em populações com frequência alélica 0,9 e diferentes coeficientes de endogamia.

Tabela 2 – Valores esperados para a variância teórica do estimador do coeficiente de endogamia pelo método da análise de variância (F_1) calculados pela expressão de Muniz, Vencovsky e Barbin (1997 a) em populações endogâmicas e diferentes tamanhos de amostra.

n	F		
	0,1	0,5	0,9
10	0,0980	0,0563	0,0036
30	0,0327	0,0188	0,0012
50	0,0196	0,0113	0,0007
100	0,0098	0,0056	0,0004
200	0,0049	0,0028	0,0002

Tabela 3 – Valores esperados para a variância teórica do estimador do coeficiente de endogamia pelo método dos momentos (F_2), calculados pela expressão de Weir (1996), em populações endogâmicas de dois alelos com diferentes combinações de frequências alélicas e tamanho de amostra.

n	p		
	0,1	0,5	0,9
10	0,0180	0,0500	0,0180
30	0,0060	0,0167	0,0060
50	0,0036	0,0100	0,0036
100	0,0018	0,0050	0,0018
200	0,0009	0,0025	0,0009

Tabela 4 – Valores esperados para a variância teórica do estimador do coeficiente de endogamia pelo método da máxima verossimilhança (F_3), calculados pela expressão de Weir (1996), em populações endogâmicas de dois alelos com diferentes combinações de frequências alélicas, coeficiente de endogamia e tamanho de amostra.

p	F	n				
		10	30	50	100	200
0,1	0,1	0,1598	0,0533	0,0320	0,0160	0,0080
	0,5	0,2083	0,0694	0,0417	0,0208	0,0101
	0,9	0,0542	0,0181	0,0108	0,0054	0,0027
0,5	0,1	0,0846	0,0282	0,0169	0,0085	0,0042
	0,5	0,0750	0,0250	0,0150	0,0075	0,0038
	0,9	0,0190	0,0063	0,0038	0,0019	0,0010
0,9	0,1	0,1598	0,0533	0,0320	0,0106	0,0080
	0,5	0,2083	0,0694	0,0417	0,0208	0,0104
	0,9	0,0542	0,0181	0,0108	0,0054	0,0027

CONCLUSÕES

Os três estimadores se mostraram tendenciosos, sendo o método da análise de variância o de menor viés, seguido pelo método dos momentos e da máxima verossimilhança.

Para amostras de tamanhos inferiores a 50, o método da análise de variância apresentou menor variância.

Os três estimadores apresentaram variâncias semelhantes quando se utilizou tamanho de amostra superior a 50 indivíduos.

Por suas propriedades, o método da análise de variância seria o indicado para a estimação do coeficiente de endogamia.

REFERÊNCIAS BIBLIOGRÁFICAS

- AYRES, K.L.; BALDING, D.J. Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. **Heredity**, Oxford, v.80, p.769-777, 1998.
- BAILEY, T.J. Testing the solubility of maximum likelihood equations in the routine application of scoring methods. **Biometrics**, Alexandria, v.7, p.268-274, 1951.
- COCKERHAM, C.C. variance of gene frequency. **Evolution**, Lancaster, v.23, p.72-84, 1969.
- FALCONER, D.S. **Introduction of quantitative genetics**. New York: The Ronald, 1964. 365p.
- FISHER, R.A. **The theory of inbreeding**. Edinburg: Oliver and Boyd, 1949. 120p.
- HARTL, D.L.; CLARK, A.G. **Principles of population genetics**. Sunderland: Sinauer Associates, 1989. 681p.
- HILL, W.G.; BABIKER, H.A.; RANFORD-CARTWRIGHT, T.; WLLIKER, D. Estimation of inbreeding coefficients from genotypic data on multiple alleles. and application to estimation of clonality in malaria parasites. **Genetics Research**, Cambridge, v.65, p.53-61, 1995.
- MOOD, A.M.; GLAYBILL, F.A.; BOES, D.C. **Introduction to theory of statistics**. 3.ed. Tokyo: McGraw-Hill Kogakusha, 1974. 564p.
- MUNIZ, J.A.; BARBIN, D.; VENCOVSKY, R. properties of estimators of the inbreeding coefficient and the rate of cross fertilization obtained from gene frequency data in a diploid population. **Brazilian Journal of Genetics**, Ribeirão Preto, v.19, n.3, p.485-491, 1996.
- MUNIZ, J.A.; BARBIN, D.; VENCOVSKY, R.; VEIGA, R.D. Teste de hipóteses sobre o coeficiente de endogamia de uma população diplóide. **Ciência e Agrotecnologia**, Lavras, v.23, n.2, p.410-420, abr./jun. 1999.
- MUNIZ, J.A.; ITO, S.C.S.; FERREIRA, D.F.; VEIGA, R.D. Teste de hipóteses sobre o coeficiente de coancestria de populações haplóides. **Pesquisa Agropecuária Brasileira**, Brasília, v.36, n.1, p.15-25, jan. 2001a.
- MUNIZ, J.A.; ITO, S.C.S.; VEIGA, R.D.; FERREIRA, D.F. Propriedades do estimador do coeficiente de coancestria com dados de frequências alélicas em populações haplóides. **Ciência e Agrotecnologia**, Lavras, v.25, n.6, p.1393-1405, nov./dez. 2001b.
- MUNIZ, J.A.; VENCOVSKY, R.; BARBIN, D. A variância do estimador do coeficiente de endogamia obtido pelo método dos momentos em uma população diplóide. **Revista de Matemática e Estatística**, São Paulo, v.15, p.131-143, 1997a.
- MUNIZ, J.A.; VENCOVSKY, R.; BARBIN, D. Estimação do coeficiente de endogamia através do método dos momentos em uma população diplóide com alelos múltiplos. **Ciência e Agrotecnologia**, Lavras, v.21, p.150-159, 1997b.
- MUNIZ, J.A.; CAMARGO, M.S.; FERREIRA, D.F.; VEIGA, R.D. Métodos de estimação do coeficiente de endogamia em uma população diplóide com alelos múltiplos. **Ciência e Agrotecnologia**, Lavras, v.32, p.93-101, 2008.
- NEIR, M.; CHESSER, R.K. Estimation of fixation indices and gene diversities. **Annual Human Genetic**, Cambridge, v.47, p.253-259, 1983.
- OLIVEIRA, R. L. de; MUNIZ, J. A.; ANDRADE, M. J. B. de; REIS, R. L. dos. Precisão experimental em ensaios com a cultura do feijão. **Ciência e Agrotecnologia**, Lavras, v.33, n.1, p. 113-119, jan./fev., 2009.
- REIS, R. L. dos; MUNIZ, J. A.; FONSECA E SILVA, F.; AQUINO, L. H. de. Estimativas de variância genética aditiva em populações selecionadas e não-selecionadas via simulação Monte Carlo utilizando o software R. **Ciência e Agrotecnologia**, Lavras, v.33, n.1, p.285-291, jan./fev., 2009.
- REYNOLDS, J.; WEIR, B.S.; COCKERHAM, C.C. Estimation of the coancestry coefficient: basis for a short-term genetic distance. **Genetics**, Baltimore, v.105, p.767-779, 1983.
- ROBERTSON, A.; HILL, W.G. Deviations of Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficient. **Genetics**, Baltimore, v.107, p.703-718, 1984.

ROUSSET, F.; RAYMOND, M. Testing heterozygote excess and deficiency. **Genetics**, Baltimore, v.140, p.1413-1419, 1995.

SILVA, R. B. V.; FERREIRA, D. F.; NOGUEIRA, D. A. Robustness of asymptotic and bootstrap tests for multivariate homogeneity of covariance matrices. **Ciência e Agrotecnologia**, Lavras, v.32, n.1, p.157-166, jan./fev., 2008.

STATISTICAL ANALYSIS SYSTEM INSTITUTE. **Software reference**. Version 6. Cary, 1990. v.1, 794p.

VENCOVSKY, R. Análise de variância de frequências alélicas. **Revista Brasileira de Genética**, Ribeirão Preto, v.15, n.1, p.53-60, 1992. Suplemento.

WEIR, B.S. **Genetic data analysis II**: methods for discrete population genetics. 2.ed. Sunderland: Sinauer Associates, 1996. 445p.

WEIR, B.S.; COCKERHAM, C.C. Estimating F-statistics for the analysis of population structure. **Evolution**, Lancaster, v.38, n.6, p.1358-1370, 1984.

WRIGHT, S. System of mating. **Genetics**, Baltimore, v.6, p.111-178, 1921.

WRIGHT, S. The interpretation of population structure by F-statistics with special regard to system of mating. **Evolution**, Lancaster, v.19, p.395-420, 1965.