

MÉTODOS DE ESTIMAÇÃO DO COEFICIENTE DE ENDOGAMIA EM UMA POPULAÇÃO DIPLÓIDE COM ALELOS MÚLTIPLOS¹

Methods of estimation of the inbreeding coefficient in a diploid population with multiple alleles

Joel Augusto Muniz², Mariele Santana Camargo³, Daniel Furtado Ferreira⁴, Ruben Delly Veiga⁴

RESUMO

Com o presente trabalho, objetivou-se avaliar as propriedades de três estimadores do coeficiente de endogamia, F , em uma população diplóide com alelos múltiplos, por meio de dados de frequências alélicas de amostras de indivíduos, obtidas em populações simuladas, por meio do SAS. Foram avaliados o estimador de F , obtido pela média das estimativas nas análises de cada alelo, o estimador considerando a análise conjunta envolvendo todos os alelos, bem como aquele por meio de análise multivariada com os três alelos proposto por Long (1986). Os resultados encontrados para a média e variância dos 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 conjunta foi menos tendencioso e apresentou menor variância, quando o coeficiente de endogamia na população era alto, enquanto que para populações com endogamia baixa a variância do estimador considerando a análise multivariada foi menor.

Termos para indexação: Frequência alélica, simulação, variação genética.

ABSTRACT

The present work evaluated the properties of three estimators of the inbreeding coefficient, F , in a diploid population with multiple alleles, using data of gene frequencies in individuals from random samples, obtained in simulate populations, through the SAS. Were evaluated the estimator of F , obtained by single and joint univariate analysis and the estimator of F obtained by multivariate analysis as proposed by Long (1986). The analysis of the means and variances of the estimators, obtained of 1000 estimates of F , calculated for each sample size, it demonstrated that the three estimators is bias. However, it was observed that the estimator obtained of univariate analysis it was less biased and it 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 it was smaller.

Index terms: Gene frequencies, simulation, genetic variation.

(Recebido em 10 de março de 2006 e aprovado em 14 de fevereiro de 2007)

INTRODUÇÃO

O conhecimento da variabilidade genética é básico para o entendimento da dinâmica evolutiva das populações. Num programa de melhoramento genético é importante conhecer bem as fontes de variabilidade, com o objetivo de se reunir num só indivíduo ou plantas, características de interesse ao melhorista. Essa variabilidade pode ser conhecida, por exemplo, quantificando-se as frequências alélicas da população, pois, a partir destas, pode-se obter estimativas de parâmetros genéticos por meio da análise de variância dos dados, e com estas estimativas descrever a estrutura genética da população. Cockerham & Weir (1983) apontam os coeficientes de endogamia e de coancestria, bem como outras medidas de identidade por

descendência dos genes, como parâmetros importantes em genética quantitativa e de populações. Estes parâmetros são úteis para informar sobre homocigosidade, deriva, endogamia e variação quantitativa.

O coeficiente de endogamia, é um parâmetro que afeta diretamente a diversidade genética de uma população tendo sido introduzido por Wright (1951), e resume de forma conveniente a estrutura de uma população. Há entretanto, alguma discordância, entre os autores que estudaram o coeficiente de endogamia, quanto a interpretação do seu valor e principalmente quanto a forma de estimá-lo, em diferentes situações.

Falconer (1964) e Hartl & Clark (1989) definem endogamia como o acasalamento entre indivíduos que são relacionados por descendência, tendo como primeiro efeito

¹Trabalho desenvolvido com o apoio do CNPq.

²Doutor, Professor Titular – Departamento de Ciências Exatas/DEX – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – Bolsista do CNPq – joamuniz@ufla.br

³Graduanda – Departamento de Ciências Exatas/DEX – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – Bolsista do CNPq/Balcão – marieleesc@hotmail.com

⁴Doutores, Professores – Departamento de Ciências Exatas/DEX – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – danielff@ufla.br, delly@ufla.br

uma mudança nas frequências genotípicas de Hardy-Weinberg devido a um aumento na frequência de genótipos homocigóticos à custa de frequência de genótipos heterocigóticos.

Wright (1965) esclarece o significado do coeficiente de endogamia e de parâmetros relacionados a pares de gametas, definindo para uma população natural hierarquicamente subdividida, os seguintes parâmetros de descrição: 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 e F_{ST} - expressa a correlação entre os gametas ao acaso dentro da subpopulação em relação aos gametas da população total. Os três parâmetros apresentam a seguinte relação: $1 - F_{IT} = (1 - F_{IS})(1 - F_{ST})$.

O seguinte modelo aleatório foi estabelecido por Cockerham (1969) para descrever a estrutura de populações, pressupondo que todas as entidades da estrutura hierárquica são obtidas por amostragem: $x_{kij} = p + a_k + b_{ki} + w_{kij}$, onde: x_{kij} representa o efeito do alelo j , no indivíduo i , do grupo k ; a_k é o efeito do grupo k , com $k = 1, 2, \dots, K$; b_{ki} é o efeito do indivíduo i dentro do grupo k , com $i = 1, 2, \dots, n_k$ e $\sum n_k = N$ e w_{kij} é o efeito do alelo j dentro do indivíduo i dentro do grupo k , $j = 1, 2$. Por meio desse modelo medidas análogas às estatísticas F de Wright, foram obtidas e a seguinte correspondência pode ser observada: $F = F_{IT}$, $q = F_{ST}$ e $f = F_{IS}$.

Vencovsky (1992) descreve a técnica de análise de variância com frequências alélicas, 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 deriva genética nas gerações passadas. São apresentados os parâmetros genéticos e suas relações com os componentes de variância associados à análise de variância das frequências alélicas. Os parâmetros são estimados por meio do método dos momentos, utilizando-se a aproximação que considera a estimativa de um quociente como sendo o quociente entre as estimativas.

A estimação do coeficiente de endogamia de uma população utilizando dados de frequências alélicas, de acordo com Vencovsky (1992) e Weir (1996) pode ser feita por meio da análise de variância em relação a uma variável binária y , que assume o valor **1** quando um determinado alelo, por exemplo A_1 de um loco, está no indivíduo e assume o valor **zero** quando este alelo está ausente e presentes os

alelos A_2, A_3, \dots, A_u . A técnica de análise é a mesma da Estatística Experimental, sendo feita a associação de um modelo aleatório conforme a estrutura apresentada pelos dados. Para uma população que tenha g locos, com u alelos ou um número variável de alelos em cada um, devem ser feitas gu ou

$$\sum_{i=1}^g u_i$$

análises de variância, respectivamente, a partir das quais se estimam os componentes de variância e os parâmetros genéticos a eles associados, devendo-se combinar as estimativas em relação a todos alelos para se ter as informações desejadas da população. Para se fazer inferência a respeito dos parâmetros genéticos envolvidos no modelo, os autores comentam que existem dificuldades pois a variável indicadora não apresenta distribuição normal e, em conseqüência, os estimadores que são dados em função de quadrados médios da análise de variância, não possuem distribuição definida.

Reynolds et al. (1983), estudando a estimação do coeficiente de coancestria q , no caso de uma população com u alelos, sugerem um estimador médio sobre todos alelos, por dois métodos. Segundo os autores, para um determinado alelo k do qual se tem o estimador de q , dado por

$$\hat{\theta}_k = \frac{a_k}{a_k + b_k}$$

em que a_k e b_k são, respectivamente, os estimadores dos componentes de variância entre populações e dentro de populações, o primeiro método consiste em se obter uma média aritmética em relação a todos alelos dada por

$$\hat{\theta}_m = \frac{1}{u} \sum_{k=1}^u \hat{\theta}_k,$$

enquanto que o segundo método baseia-se no cálculo de uma média somando-se o numerador e o denominador dos estimadores individuais, isto é,

$$\hat{\theta}_c = \frac{\sum_{k=1}^u a_k}{\sum_{k=1}^u (a_k + b_k)}$$

que corresponde a estimar q por meio de uma análise conjunta envolvendo todos os k alelos.

Muniz et al. (1997) propuseram fórmulas de estimação para o coeficiente de endogamia, obtido pelo método dos momentos na análise de variância de uma amostra de indivíduos de uma população diplóide com

alelos múltiplos, baseando-se nos métodos propostos por Reynolds et al. (1983). Os autores avaliaram as propriedades dos estimadores, observando que os mesmos são tendenciosos, sendo que o estimador obtido por meio da análise conjunta foi menos tendencioso que aquele considerando a média das estimativas de cada alelo. Observaram ainda, que os estimadores apresentam mesma variância. Resultados de simulação mostraram que as fórmulas propostas apresentam valores semelhantes e satisfatórios quando a frequência do alelo raro na população foi não inferior a 0,10; situação na qual, o estimador obtido pela média dos alelos, distanciou-se dos valores paramétricos.

A estimação de parâmetros genéticos utilizando dados de frequências alélicas com técnicas multivariadas é apresentada em trabalhos de Smouse & Neel (1977) e Smouse & Williams (1982).

Long (1986) descreve uma estrutura de população, na qual se considera um loco com Z alelos designados por A_1, A_2, \dots, A_z , com frequências populacionais p_1, p_2, \dots, p_z . A variabilidade deste loco é avaliada no contexto de um conjunto de vetores aleatórios, \mathbf{x} , cada um deles com Z - 1 elementos. O primeiro elemento do vetor é considerado 1 se o alelo é do tipo A_1 , e zero caso contrário. À segunda posição um valor 1 é atribuído se o alelo é tipo A_2 e zero caso contrário. E assim até que a posição Z - 1 seja encontrada. Dessa forma, o vetor correspondente ao alelo A_z é um vetor com Z - 1 zeros. Este modelo também estima as estatísticas F de Wright (1965), pois a matriz, Σ , de variância e de covariância pode ser subdividida nos respectivos componentes, correspondendo às várias subdivisões da população descritas pelo modelo linear proposto por Cockerham (1969), no qual, todos os efeitos são admitidos como aleatórios, independentes e apresentam matrizes de variância e covariância Σ_p, Σ_I e Σ_G .

As estimativas das matrizes, $\Sigma, \Sigma_p, \Sigma_I, \Sigma_G$, são obtidas por meio de um conjunto de Matrizes de Quadrados e de Produtos Médios (MQPM), correspondentes a cada nível hierárquico, como observado na análise de variância apresentada na Tabela 1.

Considerando $\Sigma_g = 0$, obtém-se: $\Sigma_I = F_{IT}$ e $\Sigma_p = F_{ST}$. Assim, os parâmetros associados ao modelo de Cockerham (1969) podem ser descritos pelos componentes das matrizes de variância e covariância da seguinte maneira: $(1 - F_{IT})\Sigma = \Sigma_G$; $(F_{IT} - F_{ST})\Sigma = \Sigma_I$ e $F_{ST}\Sigma = \Sigma_p$, onde: $\Sigma = \Sigma_p + \Sigma_I + \Sigma_G$.

E as estatísticas F de Wright são dadas pelas seguintes expressões:

$$\bullet F_{IS} = \left(\frac{1}{G}\right) \text{TR} \left[(\Sigma_I + \Sigma_G)^{-1/2} \Sigma_I (\Sigma_I + \Sigma_G)^{-1/2} \right]$$

$$\bullet F_{ST} = \left(\frac{1}{G}\right) \text{TR} \left[\Sigma^{-1/2} \Sigma_p \Sigma^{-1/2} \right]$$

$$\bullet F_{IT} = \left(\frac{1}{G}\right) \text{TR} \left[\Sigma^{-1/2} (\Sigma_p + \Sigma_I) \Sigma^{-1/2} \right]$$

sendo G a dimensão das matrizes Σ_p, Σ_I e Σ_G .

Os estimadores das estatísticas F de Wright são dados respectivamente por:

$$\bullet \hat{F}_{IS} = \left(\frac{1}{G}\right) \text{TR} \left[(S_I + S_G)^{-1/2} S_I (S_I + S_G)^{-1/2} \right]$$

$$\bullet \hat{F}_{ST} = \left(\frac{1}{G}\right) \text{TR} \left[S^{-1/2} S_p S^{-1/2} \right] \text{ e}$$

$$\bullet \hat{F}_{IT} = \left(\frac{1}{G}\right) \text{TR} \left[S^{-1/2} (S_p + S_I) S^{-1/2} \right].$$

sendo, S_p, S_I , e S_G os respectivamente, os estimadores das matrizes de variância e covariância de população, indivíduo e erro (Σ_p, Σ_I e Σ_G). Segundo Weir & Cockerham (1984), somente F_{IS} pode ser estimado, quando estão sendo considerados dados de frequências alélicas de uma única população.

Com o presente trabalho, objetivou-se a avaliar, por meio de simulação, as propriedades dos estimadores do coeficiente de endogamia, para uma população diplóide com alelos múltiplos, considerando a média das estimativas de cada alelo, a estimativa utilizando a análise conjunta envolvendo todos os alelos e a estimativa baseada na análise multivariada com todos os alelos.

Tabela 1 – Matrizes de Quadrados e de Produtos Médios (MQPM), correspondentes a cada nível hierárquico da população descritas pelo modelo linear proposto por Cockerham (1969).

Causas de Variação	GL	MQPM	E (MQPM)
Populações	K-1	MQPM (a)	$\Sigma_G + 2 \Sigma_I + c \Sigma_p$
Indivíduos dentro de Populações	N-K	MQPM (b)	$\Sigma_G + 2 \Sigma_p$
Alelos d. de Indivíduos d. Populações	N	MQPM (w)	Σ_G

MATERIAL E MÉTODOS

Foram desenvolvidos dois programas utilizando o *Statistical Analysis System* (SAS INSTITUTE, 1990), para simulação e estimação do coeficiente de endogamia, sendo o primeiro responsável pela realização da estimação baseada na análise univariada, ou seja, por meio da média das estimativas e da análise de variância conjunta das frequências alélicas, e o segundo pela estimação considerando a análise multivariada. Os dois programas apresentavam a mesma metodologia de simulação das populações e experimentos, pela qual foram geradas populações endogâmicas de tamanho N=1000 indivíduos com três alelos e coeficientes de endogamia variando de 0,1 no intervalo entre zero e um. As frequências alélicas (p_1, p_2, p_3) consideradas na construção da população, variaram da seguinte forma:

0,40; 0,30; 0,30 0,50; 0,25; 0,25 0,60; 0,30; 0,10 0,60; 0,39; 0,01 0,50; 0,45; 0,05
 0,50; 0,30; 0,20 0,33; 0,33; 0,34 0,50; 0,40; 0,10 0,60; 0,20; 0,20 0,50; 0,49; 0,01

Foram simuladas 110 populações, e, em cada uma delas 1000 experimentos para cada um dos tamanhos de amostras considerados (n = 10, 20, 30, 50, 100 e 200 indivíduos), totalizando-se no estudo 660 000 experimentos.

Na construção de cada uma das populações, a simulação foi feita considerando-se o intervalo entre zero e um dividido em seis segmentos proporcionais às frequências genotípicas de indivíduos $A_1A_1, A_1A_2, A_1A_3, A_2A_2, A_2A_3, A_3A_3$, conforme estrutura estabelecida para a população, sendo que, estas foram construídas pressupondo-se equilíbrio de Wright (1965). Os experimentos foram simulados utilizando-se a função RANUNI (SAS INSTITUTE, 1990) que produz valores para uma variável aleatória com distribuição uniforme no intervalo entre 0 e 1, obtendo-se assim as amostras dentro de cada população.

No caso de uma amostra de indivíduos de uma população diplóide com dois ou mais alelos, a variável indicadora de um determinado alelo tem a seguinte

descrição, segundo o modelo de Cockerham (1969): $y_{ij} = p + a_i + g_{(ji)}$, sendo: y_{ij} a frequência do alelo j dentro do indivíduo i ; p a frequência paramétrica do alelo A na população; a_i o efeito do indivíduo i , com $i = 1, 2, \dots, n$ e $g_{(ji)}$ o efeito do alelo j dentro do indivíduo i , com $j = 1, 2$. Considerando-se modelo aleatório e a metodologia sugerida por Vencovsky (1992), obtém-se o esquema apresentado na Tabela 2, para realização de análise de variância dos dados.

De acordo com a análise estimou-se o coeficiente de endogamia em cada experimento fazendo-se a média das estimativas individuais de cada um dos três alelos, dada por

$$\bar{F} = \frac{1}{u} \sum_{i=1}^u \frac{QMI_i - QMG_i}{QMI_i + QMG_i}$$

sendo u o número de alelos, QMI_i e QMG_i , respectivamente, os quadrados médios de Indivíduos e de Genes dentro de Indivíduos em cada experimento. Foi feita também a estimação do coeficiente de endogamia através da análise de variância conjunta envolvendo os três alelos. Neste caso, o coeficiente de endogamia foi estimado pela fórmula

$$\tilde{F} = \frac{\sum_{i=1}^u (QMI_i - QMG_i)}{\sum_{i=1}^u (QMI_i + QMG_i)}$$

Foram calculadas as médias dos estimadores ($\mu_{\bar{F}}$ e $\mu_{\tilde{F}}$) e suas variâncias ($Var_{\bar{F}}$ e $Var_{\tilde{F}}$), para cada amostra considerada, com base nos 1000 experimentos, nos seis tamanhos de amostra em cada uma das 110 populações simuladas. Para cada conjunto de 1000 experimentos, estimou-se também o coeficiente de endogamia segundo metodologia sugerida por Long (1986), baseando-se no esquema de análise de variância multivariada apresentado na Tabela 3, onde todos os efeitos foram considerados aleatórios, independentes, e apresentando matrizes de variância e covariância (Σ_G, Σ_I).

Tabela 2 – Análise de variância univariada, para dados de uma população diplóide, segundo Vencovsky (1992).

Causas de Variação	GL	QM	E(QM)
Entre Indivíduos	n-1	QMI	p(1-p)[(1-F) + 2F]
Alelos dentro de Indivíduos	n	QMG	p(1-p)[(1-F)]

Tabela 3 – Análise de variância multivariada, para dados de uma população diplóide, de acordo com metodologia proposta por Long (1986).

Causas de Variação	GL	MQPM	E (MQPM)
Entre Indivíduos	n-1	MQPM (I)	$\Sigma_G + 2 \Sigma_I$
Alelos dentro de Indivíduos	n	MQPM (G)	Σ_G

Com base na análise de variância, o coeficiente de endogamia pode ser obtido pela seguinte expressão:

$$\bullet \mathbf{F}^* = \left(\frac{\mathbf{1}}{\mathbf{G}} \right) \text{TR} \left[(\Sigma_I + \Sigma_G)^{-1/2} \Sigma_I (\Sigma_I + \Sigma_G)^{-1/2} \right]$$

Assim, o estimador do coeficiente de endogamia, considerando-se o enfoque multivariado é dado pela seguinte expressão:

$$\bullet \hat{\mathbf{F}}^* = \left(\frac{\mathbf{1}}{\mathbf{G}} \right) \text{TR} \left[(\mathbf{S}_I + \mathbf{S}_G)^{-1/2} \mathbf{S}_I (\mathbf{S}_I + \mathbf{S}_G)^{-1/2} \right]$$

sendo G a dimensão das matrizes \mathbf{S}_I e \mathbf{S}_G . Nas diversas populações, foram calculadas a média do estimador multivariado ($\mu_{\hat{\mathbf{F}}^*}$) e sua variância ($\text{Var}_{\hat{\mathbf{F}}^*}$), utilizando-se os resultados obtidos para os 1000 experimentos nos diversos tamanhos de amostra.

Os componentes das matrizes de variância e covariância estimadas para indivíduos (\mathbf{S}_I) e para genes dentro de indivíduos ou erro (\mathbf{S}_G), foram obtidos pela descrição detalhada dos mesmos no programa, e o valor do estimador foi calculado utilizando o procedimento **Interactive Matrix Language (IML)** do SAS.

A matriz $(\mathbf{S}_I + \mathbf{S}_G)^{-1/2}$ foi obtida por cálculo matricial da seguinte maneira:

$$(\mathbf{S}_I + \mathbf{S}_G)^{-1/2} = \mathbf{P}' \Lambda^{-1/2} \mathbf{P},$$

onde P é a matriz constituída pelos autovetores de $(\mathbf{S}_I + \mathbf{S}_G)$ e Λ é matriz diagonal dos autovalores de $(\mathbf{S}_I + \mathbf{S}_G)$. A matriz $\Lambda^{-1/2}$ foi calculada pela inversa do fator de Cholesky de Λ (SAS INSTITUTE, 1989).

As matrizes de variância e covariância envolvidas no estudo foram de dimensão 2×2 , conforme Hoffmann & Vieira (1998), uma vez que o envolvimento do terceiro alelo, levaria a uma matriz positiva não definida por ocasião da obtenção dos autovalores, uma vez, que a covariância entre os alelos poderia ser igual a 1, refletindo multicolinearidade da matriz, ou seja, a dependência linear das colunas. A estimação pela análise multivariada considerou ainda a restrição para situações onde o determinante da matriz L fatorada fosse menor ou igual a 0, não sendo possível, neste caso, a obtenção do fator de Cholesky, e também do estimador. Por essa razão, o número de experimentos considerados em cada tamanho de amostra no estudo de \mathbf{F}^* , em muitos casos, foi inferior a 1000, sendo tal fato mais evidente para pequenos tamanhos de amostra e em populações com endogamia elevada.

RESULTADOS E DISCUSSÃO

Os três estimadores do coeficiente de endogamia, F, foram comparados usando-se os valores finais fornecidos pelos dois programas de simulação. Os valores

da média e da variância para os três estimadores de F foram obtidos, para os N=1000 experimentos simulados, em cada situação envolvendo tamanho de amostra, valor do coeficiente de endogamia da população e frequências alélicas, utilizando-se o PROCEDURE UNIVARIATE (Procedimento Univarado) do SAS. Os dados foram organizados em tabelas, de forma a facilitar e orientar a visualização dos resultados obtidos para os três estimadores. Não foram apresentados todos os resultados em função do grande volume de dados. Assim, optou-se pela apresentação de duas situações, sendo uma representada por uma população que apresentou frequências semelhantes dos três alelos e outra onde observou-se frequências bastante discrepantes, indicando a ocorrência de um alelo raro. Na Tabela 4 estão apresentados os resultados obtidos para os três estimadores em populações com três alelos de mesma frequência e diferentes coeficientes de endogamia. De acordo com os resultados apresentados na Tabela 4, os três estimadores mostraram-se tendenciosos e com variâncias semelhantes, principalmente nos casos onde as populações apresentaram endogamia baixa.

Para tamanho de amostra inferior a 30, o estimador usando a análise multivariada \mathbf{F}^* foi, entre os três, o menos tendencioso quando se considerou endogamia inferior a 0,3 na população, entretanto, o viés deste estimador foi o mais elevado para populações com endogamia alta. Com o aumento do tamanho da amostra, o viés dos três estimadores tenderam a ser semelhantes. Para populações com baixa endogamia, as variâncias dos três estimadores, foram semelhantes. Observou-se diferenças entre as variâncias dos três estimadores, para populações com endogamia superior a 0,6, quando se trabalhou com amostras menores que 50 indivíduos. Nestes casos, a variância do estimador considerando a análise conjunta ($\tilde{\mathbf{F}}$), foi menor do que as variâncias dos demais estimadores.

Os resultados, em geral, comprovaram que $\tilde{\mathbf{F}}$, apresentou resultados de média e variância mais consistentes, quando os três alelos da população apresentaram frequências semelhantes e para pequenos tamanhos de amostras. Na Tabela 5 estão apresentados os resultados obtidos para os três estimadores em populações com três alelos, um deles de baixa frequência e diferentes coeficientes de endogamia. No caso de populações com um alelo raro, observou-se que os três estimadores se mostraram mais tendenciosos quando comparados aos valores obtidos em populações com frequências alélicas semelhantes. O estimador que utiliza a média das estimativas, ($\bar{\mathbf{F}}$), superestimou o valor paramétrico de F para populações com endogamia inferior a 0,9, concordando com os resultados obtidos por Muniz et al. (1997). O estimador considerando o enfoque multivariado, \mathbf{F}^* , e o estimador usando a análise

Tabela 4 – Média e variância dos estimadores do coeficiente de endogamia considerando diferentes tamanhos de amostras e N=1000 experimentos simulados em populações com frequências alélicas 1/3, 1/3, 1/3 e diferentes coeficientes de endogamia (F).

Frequências Alélicas: 1/3, 1/3, 1/3								
Amostras (n)	F	N	\bar{X}_{F^*}	Var_{F^*}	$\bar{X}_{\bar{F}}$	$\text{Var}_{\bar{F}}$	$\bar{X}_{\tilde{F}}$	$\text{Var}_{\tilde{F}}$
10	0,0	1000	-0,00694	0,04816	-0,01308	0,05034	-0,01301	0,05297
20		1000	-0,00248	0,02742	-0,00694	0,02384	-0,00631	0,02430
30		1000	-0,00397	0,01696	-0,00493	0,01579	-0,00434	0,01610
50		1000	-0,00357	0,01037	-0,00076	0,00986	-0,00089	0,00994
100		1000	-0,00305	0,00473	-0,00565	0,00497	-0,00563	0,00498
200		1000	-0,00569	0,00266	-0,00216	0,00249	-0,00220	0,00250
10	0,3	996	0,28130	0,05753	0,27779	0,05895	0,28710	0,05984
20		1000	0,29338	0,02759	0,28515	0,02822	0,28974	0,02852
30		1000	0,28868	0,01952	0,29602	0,01765	0,29886	0,01782
50		1000	0,29354	0,01295	0,30028	0,01173	0,30128	0,01179
100		1000	0,29452	0,00537	0,29503	0,00547	0,29564	0,00545
200		1000	0,29605	0,00267	0,29628	0,00279	0,29656	0,00279
10	0,5	993	0,47483	0,05650	0,48439	0,05223	0,49859	0,05017
20		1000	0,48670	0,02599	0,48957	0,02736	0,49501	0,02688
30		1000	0,48693	0,01839	0,49165	0,01612	0,49527	0,01616
50		1000	0,49505	0,01131	0,49912	0,00979	0,50036	0,00974
100		1000	0,49497	0,00507	0,49672	0,00516	0,49748	0,00515
200		1000	0,49535	0,00236	0,49734	0,00251	0,49767	0,00250
10	0,7	986	0,67945	0,04748	0,68226	0,04233	0,69508	0,03880
20		1000	0,68593	0,02018	0,68853	0,01997	0,69333	0,01909
30		1000	0,69190	0,01323	0,69405	0,01212	0,69710	0,01201
50		1000	0,69220	0,00805	0,69831	0,00692	0,69961	0,00688
100		1000	0,69503	0,00370	0,69653	0,00353	0,69712	0,00352
200		1000	0,69468	0,00174	0,69722	0,00187	0,69752	0,00187
10	0,9	967	0,88302	0,02384	0,88890	0,01986	0,89614	0,01606
20		1000	0,89363	0,00843	0,89353	0,00747	0,89560	0,00703
30		1000	0,89766	0,00541	0,89539	0,00501	0,89678	0,00485
50		1000	0,89267	0,00300	0,89452	0,00279	0,89513	0,00275
100		1000	0,89496	0,00158	0,89588	0,00150	0,89622	0,00149
200		1000	0,89520	0,00070	0,89640	0,00072	0,89653	0,00072

F^* - análise multivariada; \bar{F} - média dos alelos; \tilde{F} - análise conjunta

conjunta, \tilde{F} , subestimaram o valor paramétrico, verificando-se para o estimador \tilde{F} , menor tendência. Com o aumento da endogamia, observou-se a redução na variância dos três estimadores, a exceção de \tilde{F} , que apresentou menor variância quando a endogamia da população foi inferior a 0,6, para todos tamanhos de amostra.

As Figuras 1, 2, 3 e 4, ilustram o comportamento da média e da variância dos três estimadores, no caso de

populações com endogamia total e com baixa endogamia, considerando-se os três alelos com frequências semelhantes e com a ocorrência de um alelo raro. Por meio das figuras observa-se que com o aumento do tamanho da amostra em populações onde os alelos apresentavam frequências semelhantes, a variabilidade do estimador de F^* pela análise multivariada é maior em populações com endogamia total, do que em populações, onde a endogamia

Tabela 5 – Média e variância dos estimadores do coeficiente de endogamia considerando diferentes tamanhos de amostras e N=1000 experimentos simulados em população de frequências alélicas 0,50; 0,49; 0,01 e diferentes coeficientes de endogamia (F).

Frequências Alélicas: 0,50; 0,49; 0,01								
Amostras (n)	F	N	\bar{X}_{F^*}	Var_{F^*}	$\bar{X}_{\bar{F}}$	$\text{Var}_{\bar{F}}$	$\bar{X}_{\tilde{F}}$	$\text{Var}_{\tilde{F}}$
10	0,0	182	-0,00835	0,02706	0,26901	0,06145	-0,01669	0,09900
20		332	-0,00349	0,01121	0,23434	0,04603	-0,00038	0,04861
30		393	-0,00099	0,00889	0,19949	0,04272	-0,00769	0,03299
50		613	-0,00288	0,00517	0,13754	0,03622	0,00217	0,01840
100		832	-0,00238	0,00242	0,04869	0,01894	-0,00104	0,00949
200		975	-0,00472	0,00128	0,00328	0,00451	-0,00242	0,00485
10	0,3	137	0,14217	0,02749	0,47196	0,05876	0,27194	0,09573
20		265	0,13958	0,01188	0,44347	0,04403	0,29892	0,04308
30		340	0,14504	0,00826	0,41409	0,04107	0,28796	0,03108
50		562	0,14629	0,00453	0,34654	0,03776	0,28870	0,01765
100		804	0,14574	0,00239	0,27010	0,02307	0,29353	0,00895
200		951	0,14602	0,00116	0,21539	0,00834	0,29028	0,00412
10	0,5	128	0,24470	0,03615	0,61300	0,04668	0,46413	0,07950
20		247	0,27829	0,02983	0,59612	0,03496	0,48881	0,03698
30		300	0,27728	0,01817	0,56414	0,03518	0,48371	0,02522
50		502	0,28180	0,01830	0,51528	0,03491	0,48829	0,01477
100		726	0,29661	0,01875	0,43154	0,02737	0,48898	0,00740
200		925	0,29250	0,01258	0,37737	0,01201	0,48974	0,00376
10	0,7	104	0,40639	0,05338	0,75541	0,03292	0,66693	0,05400
20		209	0,49149	0,06182	0,74111	0,02973	0,68598	0,02618
30		268	0,49581	0,05589	0,72921	0,02664	0,68857	0,01694
50		440	0,49682	0,04976	0,69284	0,02818	0,68676	0,01018
100		661	0,49942	0,04420	0,63450	0,02823	0,68916	0,00535
200		892	0,51255	0,03489	0,59366	0,01925	0,68891	0,00251
10	0,9	79	0,67173	0,08366	0,91227	0,01479	0,88532	0,02148
20		174	0,74121	0,06228	0,90440	0,01374	0,89275	0,01055
30		221	0,72447	0,06126	0,89972	0,01324	0,89341	0,00694
50		374	0,75128	0,05486	0,88190	0,01523	0,89184	0,00406
100		596	0,74450	0,05139	0,85040	0,01926	0,89205	0,00214
200		833	0,75803	0,03871	0,82281	0,01709	0,89278	0,00099

F^* - análise multivariada; \bar{F} - média dos alelos; \tilde{F} - análise conjunta

é baixa. No caso de populações com endogamia total as variâncias dos três estimadores são semelhantes e caem visivelmente quando o número de indivíduos amostrados é igual ou superior a 20 (Figuras 1 e 2). Para populações com um dos alelos raros (Figuras 3 e 4), observa-se que quando a endogamia foi baixa, os vieses dos estimador usando análise multivariada (F^*) e a análise conjunta (\tilde{F}) não foram sensíveis ao aumento do tamanho da amostra. O estimador que utiliza a média das estimativas (\bar{F}) de cada alelo teve

viés reduzido com o aumento do tamanho da amostra. No caso de populações com endogamia total, praticamente os três estimadores se mostraram insensíveis ao aumento do tamanho da amostra em relação à redução do viés. foi o menos tendencioso. Nas Figuras 3 e 4, verifica-se que a variância do estimador usando a análise multivariada é menor que a dos outros dois estimadores, em populações com endogamia baixa, sendo esta variância maior em populações com endogamia total.

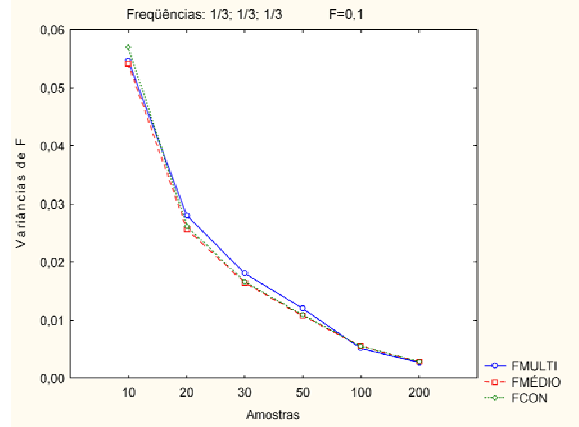
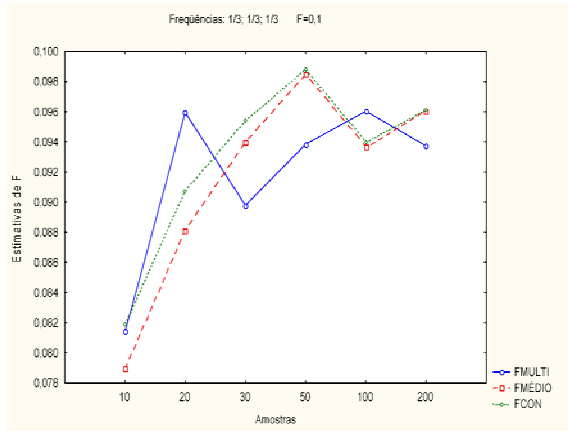


Figura 1 – Média e variância dos estimadores de F, em população com frequências alélicas de 1/3, 1/3 e 1/3 considerando endogamia de 0,1 e amostras de tamanhos diferentes.

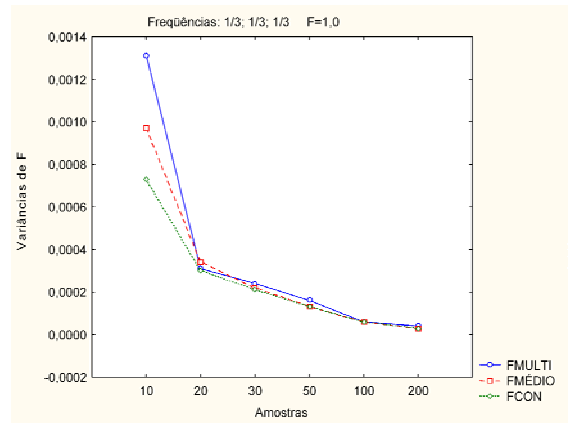
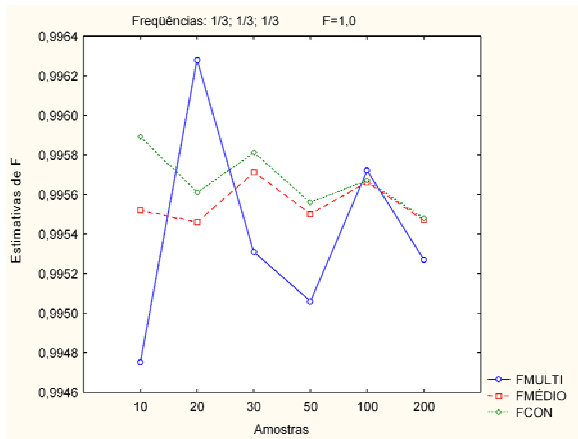


Figura 2 – Média e variância dos estimadores de F, em população com frequências alélicas de 1/3, 1/3, 1/3 considerando endogamia de 1,0 e amostras de tamanhos diferentes.

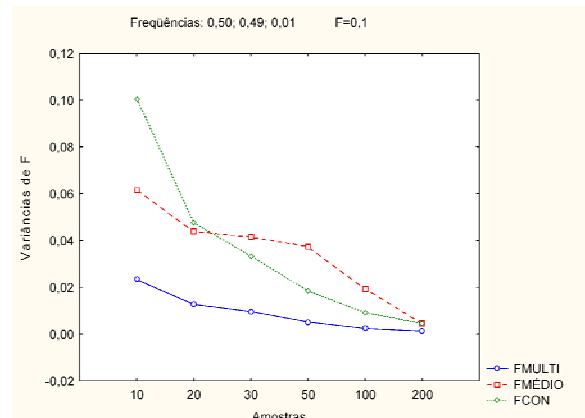
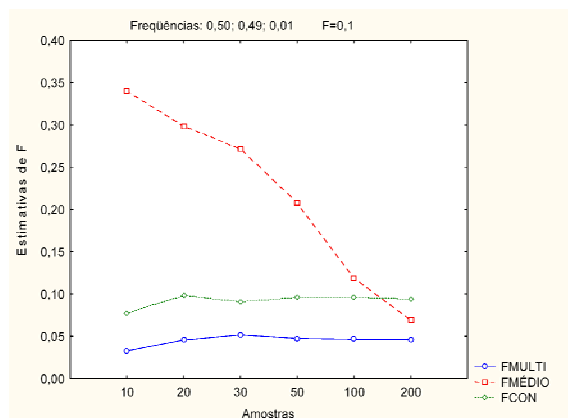


Figura 3 – Média e variância dos estimadores de F, em população com frequências alélicas de 0,50; 0,49 e 0,01 considerando endogamias de 0,1 e amostras de tamanhos diferentes.

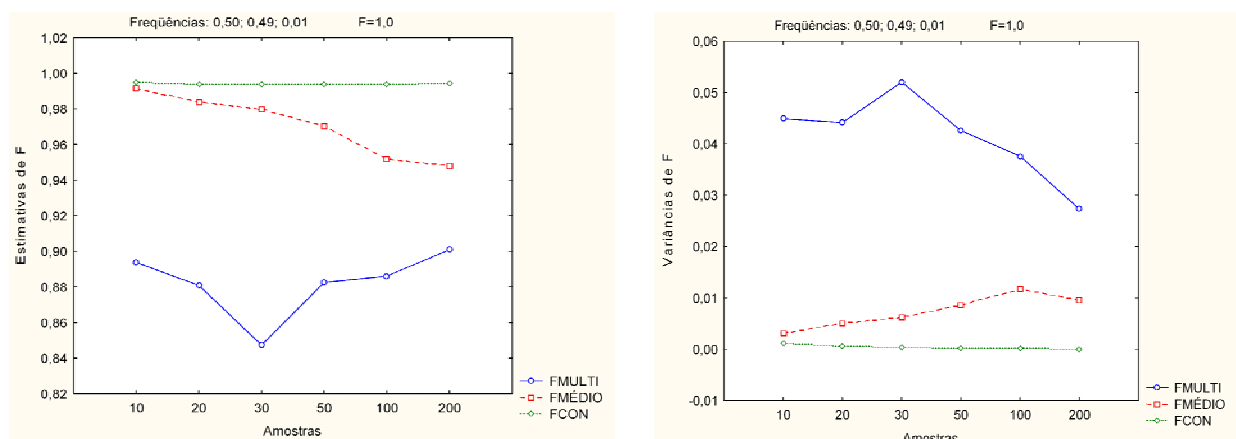


Figura 4 – Média e variância dos estimadores de F, em população com frequências alélicas de 0,50; 0,49 e 0,01, considerando endogamia de 1,0 e amostras de tamanhos diferentes.

CONCLUSÕES

Os três estimadores foram tendenciosos, sendo o viés maior para populações com ocorrência de um alelo raro e pequeno tamanho de amostra. Quando as populações apresentaram frequências alélicas iguais, observou-se valores semelhantes para as variâncias e para as tendências dos três estimadores. Na presença de um alelo raro, o viés do estimador de F por meio da análise multivariada aumentou em relação aos demais, mas apresentou menor variância em populações com endogamia inferior a 0,5. O estimador de F por meio da análise conjunta apresentou melhores propriedades que os demais, especialmente para pequenos tamanhos de amostra.

REFERÊNCIAS BIBLIOGRÁFICAS

COCKERHAM, C. C. Variance of gene frequency. *Evolution*, Lancaster, v. 23, p. 72-84, 1969.

COCKERHAM, C. C.; WEIR, B. S. Variance of actual inbreeding. *Theoretical Population Biology*, New York, v. 23, p. 85-109, 1983.

FALCONER, D. S. *Introduction of quantitative genetics*. New York: The Ronald, 1964. 365 p.

HARTL, D. L.; CLARK, A. G. *Principles of population genetics*. Sunderland: Sinauer Associates, 1989. 681 p.

HOFFMANN, R.; VIEIRA, S. *Análise de regressão: uma introdução à econometria*. 3. ed. São Paulo: Hucitec, 1998. 379 p.

LONG, J. C. The allelic correlation structure of Gainj and Kalam speaking people: I. the estimation and interpretation of Wright's F-statistics. *Genetics*, Baltimore, v. 112, p. 629-647, 1986.

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, 1997.

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.

SMOUSE, P. E.; NEEL, J. V. Multivariate analysis of genetic disequilibrium in the Yanomama. *Genetics*, Baltimore, v. 85, p. 733-752, 1977.

SMOUSE, P. E.; WILLIAMS, R. C. Multivariate analysis of HLA: disease associations. *Biometrics*, Washington, v. 38, p. 757-768, 1982.

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

STATISTICAL ANALYSIS SYSTEM INSTITUTE. *Software: usage and reference*. Version 6. Cary, 1989. 501 p.

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

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

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. The genetical structure of populations. **Annual Eugen.**, [S.l.], v. 15, p. 323-354, 1951.

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