

MÉTODOS PARA MEDIR DESVIOS DO EQUILÍBRIO DE HARDY-WEINBERG ATRAVÉS DE MEDIDAS DE ENDOCRUZAMENTO

Cibele Queiroz da-SILVA¹
Talita ARMBORST²

- RESUMO: Uma das principais medidas de discrepância do Equilíbrio de Hardy-Weinberg (EHW) é a medida de endocruzamento, f . Ela mede o déficit de heterozigosidade na população. Neste trabalho, três métodos para estimar f foram utilizados: o Bayesiano de Ayres e Balding (AB)(1998), e os clássicos de Nei e Chesser(NC)(1983) e de Robertson e Hill(RH)(1984). Analisou-se dados simulados através de frequências genotípicas considerando-se tamanhos amostrais de 50, 200 e 1.000 observações e 2, 6 e 15 alelos, sendo $n = 50$ o caso mais realístico. A comparação dos métodos foi feita a partir da construção de intervalos de credibilidade (via MCMC) e de confiança, sendo o último encontrado pelo método Bootstrap. Os resultados mostraram que o método de Ayres e Balding (1998) é, de fato, eficiente no que diz respeito à produção de estimativas de f que respeitem os limites teóricos do parâmetro. Além disso, o método de AB tem a vantagem de permitir que estimativas das frequências alélicas sejam calculadas. Porém, além de demandar cálculos muito mais complexos do que os demais, o método de Ayres e Balding (1998) pode produzir valores estimados de f com vícios apreciáveis, sendo de magnitude muito superior àqueles produzidos pelos demais métodos.
- PALAVRAS-CHAVE: Genética; medida de endocruzamento; MCMC; inferência clássica e Bayesiana.

1 Introdução

A *Genética de Populações* é um ramo da Genética que descreve, em termos matemáticos, as consequências da herança mendeliana, sob um aspecto

¹Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília – UnB, CEP: 70910-900, Brasília, Brasil, E-mail: *cibelegs@unb.br*.

²Inep/MEC - Esplanada dos Ministérios, Bloco L - Anexos II, sala 422 - CEP:70047-900 - Brasília - DF, Brasil, E-mail: *tarmborst@yahoo.com.br*

populacional, considerando um grupo de organismos que cruzam entre si e compartilham um *conjunto gênico* (*gene pool*) comum (Gardner, 1977). O conjunto gênico é o total de informações genéticas que possuem os membros férteis de uma população de organismos que se reproduzem sexualmente. Fatores evolutivos como a seleção, a migração e a mutação tendem a alterar as frequências gênicas, produzindo assim mudanças evolutivas na população.

A Lei de Hardy-Weinberg é de fundamental importância no desenvolvimento do estudo da Genética de Populações. O Equilíbrio de Hardy-Weinberg (EHW) afirma que, em uma grande população na qual os cruzamentos ocorrem de maneira aleatória, e na ausência de fatores evolutivos, as frequências alélicas e as frequências genotípicas permanecem constantes, de geração para geração. No caso de dois alelos, Hardy e Weinberg demonstraram que, quando os cruzamentos ocorrem de forma aleatória, as frequências alélicas e genotípicas seguem uma distribuição *binomial* nas populações de organismos diplóides ($2n$ conjuntos gênicos). Nesse caso, os alelos A_1 e A_2 possuem frequências p_{A_1} e $p_{A_2} = 1 - p_{A_1}$, respectivamente. No binômio de Newton associado a $(p_{A_1} + p_{A_2})^2 = 1$ ou $p_{A_1}^2 + 2p_{A_1}p_{A_2} + p_{A_2}^2 = 1$, cada termo do binômio está associado a um dos três genótipos possíveis (A_1A_1 , A_1A_2 , A_2A_2) que ocorrem na população nas seguintes proporções:

- $P_{A_1A_1} = p_{A_1}^2$: proporção do genótipo homozigoto A_1 (A_1A_1)
- $P_{A_1A_2} = 2p_{A_1}p_{A_2}$: proporção do genótipo heterozigoto (A_1A_2)
- $P_{A_2A_2} = p_{A_2}^2$: proporção do genótipo homozigoto A_2 (A_2A_2)

Tal conjunto de proporções alélicas é conhecido como proporções de Hardy-Weinberg e o denotaremos como *Modelo de Hardy-Weinberg* (HW). O estudo de violações à lei de Hardy-Weinberg é importante, uma vez que muitos outros resultados e teorias na Genética são baseadas nesse pressuposto.

Uma das medidas utilizadas para quantificar desvios do equilíbrio de Hardy-Weinberg é o coeficiente de endogamia f , conhecido também como *medida de endocruzamento*. Tal quantidade mede a proporção em que a endogamia ocorrida em dada população, que é o cruzamento entre parentes nesta população, reduz a heterozigosidade. Assim, quanto menor a quantidade de genótipos heterozigotos de uma população, maior o grau de parentesco entre os indivíduos em sucessivas gerações. Por causa dessa diminuição na heterozigosidade, os alelos recessivos se expressam mais vezes através das gerações. Isso pode ser vantajoso quando o caráter envolvido neste alelo torna o indivíduo mais bem adaptado. Porém, pode ser prejudicial se o caráter envolver doença ou, por exemplo, a deterioração da qualidade de algum fator orgânico que permita a saudável reprodução da espécie.

Uma forma de estimar a medida de endocruzamento é pelo método Bayesiano de Ayres e Balding (1998) baseado em MCMC. Outra forma se dá pelos métodos clássicos de Nei e Chesser (1983) e Robertson e Hill (1984).

1.1 Objetivos

O objetivo principal deste artigo é a comparação das características inferenciais da medida de endocruzamento, f , estimada pelos métodos clássicos e pelos métodos Bayesianos. Neste trabalho, as frequências alélicas são estimadas no caso de um *loco* com dois ou mais alelos, quando a causa principal de violações no EHW é devida à endogamia. Para tanto, foram utilizados o procedimento Bayesiano proposto por Ayres e Balding (1998) e os procedimentos clássicos de Nei e Chesser (1983) e de Robertson e Hill (1984). A motivação para o uso da Inferência Bayesiana na estimação do coeficiente de endogamia está descrita no início da Seção 4. Os resultados das inferências obtidas com a aplicação das abordagens clássica e Bayesiana foram comparados à luz das vantagens e desvantagens oferecidas por cada técnica na solução do problema em estudo.

Para a geração dos dados simulados utilizamos o modelo endogâmico descrito na Seção 2. Vários cenários envolvendo diferenças entre o número de alelos, número de indivíduos e valor populacional da medida de endocruzamento, f , foram analisados. Maiores detalhes sobre a geração dos dados simulados estão descritos na Seção 5.

2 Proporções genotípicas sob o modelo endogâmico

Considere um dado *loco* em estudo e seja p_i a proporção populacional do alelo i , $i = 1, \dots, k$. As proporções genotípicas homozigotas (P_{ii}) e heterozigotas (P_{ij}) para o caso de alelos múltiplos, sob o *modelo endogâmico* (Ayres e Balding, 1998), são:

$$\begin{cases} P_{ii} = p_i(f + (1-f)p_i) \\ P_{ij} = 2p_i p_j(1-f). \end{cases} \quad (1)$$

Considerando que $p_{min} = \min\{p_1, \dots, p_k\}$, pode-se mostrar que os *limites* de f são dados por:

$$-\frac{p_{min}}{1-p_{min}} \leq f \leq 1.$$

Note que quando $f = 0$, as proporções genotípicas seguem a lei de Hardy-Weinberg. Quando $f = 1$, isto indica que as proporções genotípicas dos heterozigotos são iguais a zero.

Segundo Murray (1996), quando há um excesso de homozigotos e, conseqüentemente, a medida de endocruzamento é positiva, isso se deve, entre outras possibilidades, a que (1) o *loco* em estudo esteja sob seleção; (2) o endocruzamento seja prática comum na população. Quando há um excesso de heterozigotos e, conseqüentemente, a medida de endocruzamento é negativa, isso pode ser devido a dois fatores principais: (1) à presença de alelos deletérios ou genes letais; (2) à predominância de cruzamentos exogâmicos (preferência por cruzamentos não-endogâmicos).

3 Estimação freqüentista da medida de endocruzamento

Há diversas maneiras de se estimar a medida de endocruzamento f . Em particular, há duas medidas freqüentistas que foram construídas levando-se em conta a presença de grupos na população, a de Nei e Chesser (1983) e a de Robertson e Hill (1984). Na maioria das vezes, os referidos grupos são estabelecidos de acordo com cruzamentos preferenciais devidos a alguma característica. Por exemplo, os indivíduos tendem a se cruzar com aqueles mais próximos ou com características em comum como estatura, origem etc.

Considere um *loco* com k alelos. A freqüência do alelo i , n_i , é obtida ao adicionarmos as freqüências genotípicas heterozigotas que envolvem o alelo i , n_{ij} , $i \neq j$ e duas vezes a freqüência genotípica homozigota, n_{ii} , isto é:

$$n_i = 2n_{ii} + \sum_{i \neq j} n_{ij}.$$

Seja n o tamanho da amostra aleatória de indivíduos selecionados na população de interesse, então

$$n = \sum_{i=1}^k n_{ii} + \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_{ij}.$$

De acordo com a notação introduzida no início da Seção 2, o vetor de freqüências genotípicas $(n_{11}, \dots, n_{ij}, \dots, n_{kk})$ segue distribuição multinomial com vetor de parâmetros $(n; P_{11}, \dots, P_{ij}, \dots, P_{kk})$, para $i \leq j \leq k$; $i = 1, \dots, k$.

As medidas de endocruzamento de Nei e Chesser (1983), \hat{f}_{nc} , e de Robertson e Hill (1984), \hat{f}_{rh} , são dadas a seguir.

$$\hat{f}_{nc} = \frac{\sum_i (\hat{P}_{ii} - \hat{p}_i^2) + (1 - \sum_i \hat{P}_{ii}) / 2n}{(1 - \sum_i \hat{p}_i^2) - ((1 - \sum_i \hat{P}_{ii}) / 2n)}; \quad (2)$$

$$\hat{f}_{rh} = \frac{1}{k-1} \sum_i \frac{2(2n-1)n_{ii} - n_i(n_i-1)}{2n_i(n-1)}. \quad (3)$$

Apesar dos estimadores \hat{f}_{nc} e \hat{f}_{rh} serem extremamente simples de calcular, tais medidas apresentam as seguintes limitações:

1. não utilizam, de forma explícita, o modelo de endocruzamento dado pela equação (1);
2. as estimativas de endocruzamento podem cair fora do intervalo onde f está definida;
3. as variâncias de \hat{f}_{nc} e \hat{f}_{rh} não possuem formas explícitas.

Considerando um dado *loco* com k alelos, é possível derivar um estimador de f pelo modelo endogâmico dado pela expressão (1). Considerando o referido modelo multinomial para frequências genóticas, a verossimilhança associada é dada por

$$L(n_{11}, \dots, n_{ij}, \dots, n_{kk} | f, \{p_i\}) = c \prod_{i=1}^k (fp_i + (1-f)p_i^2)^{n_{ii}} \prod_{j=i+1}^k (2p_i p_j (1-f))^{n_{ij}} \quad (4)$$

onde c representa o termo constante (com relação aos parâmetros p_i 's e f) de uma distribuição multinomial, sendo escrito como uma função dos n_{ij} 's.

No caso de dois alelos, o EMV (Estimador de Máxima Verossimilhança) da medida de endocruzamento é dado por

$$\hat{f}_{EMV} = 1 - \frac{2n_{12}n}{(2n_{11} + n_{12})(n_{12} + 2n_{22})}, \quad (5)$$

onde, $n_{11} = n_{A_1A_1}$, $n_{12} = n_{A_1A_2}$, $n_{22} = n_{A_2A_2}$. Quando há mais de dois alelos, faz-se necessária a utilização de métodos numéricos. Para maiores informações, ver Robertson e Hill (1984).

Quanto a estimação da incerteza de \hat{f}_{nc} e \hat{f}_{rh} , como mencionado no item (3) desta seção, um dos problemas do uso de \hat{f}_{nc} e \hat{f}_{rh} para estimar o grau de endocruzamento em uma população está na ausência de formas explícitas para o cálculo de medidas de acurácia de \hat{f}_{nc} e \hat{f}_{rh} , tais como variância e intervalos de confiança. Para contornar esse problema, faz-se necessário o uso de métodos computacionais tais como o Bootstrap (Efron, 1979; Efron e Tibshirani, 1993).

4 Estimação Bayesiana da medida de endocruzamento

Motivados pelos problemas descritos nos itens (1) a (3) da seção anterior, Ayres e Balding (1998) propuseram técnicas Bayesianas para medir desvios do modelo de HW (vide Seção 1) pela estimação da medida de endocruzamento f . Nosso interesse, neste artigo, é aplicar tais técnicas a dados reais e simulados e avaliar a qualidade das estimativas.

Quando há vários parâmetros no modelo ou os cálculos são muito sofisticados para se encontrar o estimador Bayesiano de maneira analítica, utiliza-se os métodos de Cadeia de Markov via Monte Carlo (MCMC). Esses métodos, apesar de serem computacionalmente intensivos, permitem a análise gráfica dos resultados, uma vez que se pode construir histogramas representando a densidade aproximada de uma distribuição *a posteriori*, e, portanto, uma direta interpretação.

Para obter estimativas das proporções alélicas a partir da *distribuição limite*, que é a distribuição *a posteriori* (*distribuição-alvo*), Ayres e Balding (1998) propuseram um esquema amostral utilizando o *algoritmo de Metropolis-Hastings dentro de Gibbs* (Metropolis et al., 1953; Hastings, 1970) para gerar amostras a partir dessa distribuição limite. Neste algoritmo, cada ciclo de atualização do MCMC, envolve a amostragem das proporções alélicas e da medida de endocruzamento.

As proporções alélicas amostrais são geradas segundo um procedimento em que apenas dois dos parâmetros relativos às proporções alélicas são atualizados em cada ciclo do MCMC. O objetivo de tal prática é garantir que, em cada ciclo de atualização, as proporções alélicas estimadas somem 1.

Esse ciclo é composto de dois blocos, o primeiro, **Bloco I**, consiste em amostrar duas proporções alélicas (escolhidas ao acaso), e o segundo, **Bloco II**, consiste em amostrar a medida de endocruzamento f . Assim, no Bloco I, f é fixo, e somente o vetor das proporções alélicas estimadas é atualizado. No Bloco II, acontece o inverso, as proporções alélicas estimadas são fixas e o valor de f é atualizado.

4.1 Algoritmo para estimar as proporções alélicas

O algoritmo para amostrar as duas proporções alélicas em cada ciclo é sumarizado a seguir:

Bloco I:

1. Escolha, ao acaso, dois entre os k alelos, digamos u e v .
2. Obtenha as proporções atuais dos alelos u e v , p_u e p_v , respectivamente.
3. Gere um *valor candidato*, p'_u , da distribuição geradora de candidatos q , que é uma *uniforme* com parâmetros:

$$\max(0, p_u - \varepsilon_p) \text{ e } \min(p_u + \varepsilon_p, p_u + p_v),$$

O valor do erro das proporções alélicas ε_p é escolhido de modo a se obter a taxa de aceitação (ou seja, contabiliza o número de vezes em que p'_u foi aceito, pelo algoritmo Metropolis, ao longo dos ciclos MCMC) entre 20% e 50% (Gilks; Richardson e Spiegelhater, 1996). Assim, ε_p não pode ser muito grande, pois nesse caso a cadeia não irá se mover.

4. Calcule a proporção do alelo candidato v :

$$p'_v = p_u + p_v - p'_u.$$

5. Defina a probabilidade de aceitação $\alpha(p_u, p'_u)$ como

$$\alpha(p_u, p'_u) = \min\left(\frac{g(p'_u)q(p_u|p'_u)}{g(p_u)q(p'_u|p_u)}, 1\right),$$

onde $g(p_u)$ e $g(p'_u)$ representam a distribuição condicional completa *a posteriori* de p_u e de p'_u respectivamente.

6. Gere um valor a de uma *uniforme* $(0, 1)$. Aceite p'_u se $a \leq \alpha(p_u, p'_u)$. Caso contrário, faça $p'_u = p_u$ e $p'_v = p_v$.

Os elementos descritos no algoritmo são detalhados a seguir. Para encontrar a distribuição condicional completa *a posteriori*, $g(\cdot)$, mencionada no item (5), é necessário utilizar o Teorema de Bayes:

$$\pi(p_i|f, \{n_{11}, \dots, n_{kk}\}) \propto L(f, \{p_i\})\pi(p_i).$$

Considerando uma *priori uniforme*(0,1) para a probabilidade do alelo u , p_u , temos:

$$\pi(p_u) = \begin{cases} 1, & \text{se } 0 < p_u < 1 \\ 0, & \text{caso contrário} \end{cases}$$

com a restrição de que $\sum_{i=1}^k p_i = 1$.

Nesta proposta de Ayres e Balding (1998), como se observa, não será utilizada como distribuição *a priori* a distribuição *Dirichlet* (ver Gelman et al., 2003). Desta forma, evita-se o problema da definição dos hiperparâmetros da *priori*. Também, como pode ser visto na próxima seção, tem-se mais flexibilidade na expressão do relacionamento entre o coeficiente de endocruzamento f e as proporções alélicas p_i 's. Como descrito na Seção 3, a verossimilhança relacionando o modelo endogâmico é dada pela expressão (4). Logo, a distribuição condicional completa *a posteriori*, de p_u é dada por:

$$\begin{aligned} g(p_u) &= \pi(p_u|f, \{n_{11}, \dots, n_{kk}\}) \propto L(f, \{p_i\})\pi(p_u) \\ &= \prod_{i=1}^k \left[(fp_i + (1-f)p_i^2)^{n_{ii}} \prod_{j=i+1}^k (2p_i p_j (1-f))^{n_{ij}} \right] \times I_{(0,1)}(p_u). \end{aligned} \quad (6)$$

Para encontrar a probabilidade de aceitação, é necessário calcular o termo dado no item (5), $\frac{g(p'_u)g(p_u|p'_u)}{g(p_u)g(p'_u|p_u)}$, que denominaremos B_p . Utilizando a equação (6), obtemos

$$\begin{aligned} B_p &= \frac{\prod_{i=1}^k \left[(fp'_i + (1-f)(p'_i)^2)^{n_{ii}} \prod_{j=i+1}^k (2p'_i p'_j (1-f))^{n_{ij}} \right] \times I_{(0,1)}(p'_u)}{\prod_{i=1}^k \left[(fp_i + (1-f)p_i^2)^{n_{ii}} \prod_{j=i+1}^k (2p_i p_j (1-f))^{n_{ij}} \right] \times I_{(0,1)}(p_u)} \\ &\quad \times \frac{I_{(LI',LS')}(p'_u)}{I_{(LI,LS)}(p_u)}, \end{aligned} \quad (7)$$

onde:

- $\{p_i\}$: é o vetor de proporções alélicas contendo p_u e p_v ;
- $\{p'_i\}$: é o vetor de proporções alélicas contendo p'_u e p'_v , em vez de p_u ;

- $LI = \max(0, p_u - \varepsilon_p)$ e $LI' = \max(0, p'_u - \varepsilon_p)$: limites inferiores da distribuição geradora de candidatos q quando se considera p_u ou p'_u ;
- $LS = \min(p_u + \varepsilon_p, p_u + p_v)$ e $LS' = \min(p'_u + \varepsilon_p, p'_u + p'_v)$: limites superiores da distribuição geradora de candidatos q quando se considera p_u e p_v ou p'_u e p'_v .

Supondo que $p_u \in (0, 1) \cap (LI, LS)$ e $p'_u \in (0, 1) \cap (LI', LS')$, é possível fazer uma transformação logarítmica em B_p , denominada B'_p , para simplificar os cálculos:

$$B'_p = \sum_{i=1}^k n_{ii} \ln (fp'_i + (1-f)(p'_i)^2) + \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_{ij} \ln (2p'_i p'_j (1-f)) - \sum_{i=1}^k n_{ii} \ln (fp_i + (1-f)p_i^2) - \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_{ij} \ln (2p_i p_j (1-f)). \quad (8)$$

É importante ressaltar que cada termo de $\ln(\cdot)$ na fórmula de B'_p é uma proporção genotípica. Logo, o argumento deve estar entre 0 e 1, e deve-se considerar essa restrição no problema. Com base na expressão (8), a probabilidade de aceitação é dada por

$$\alpha(p_u, p'_u) = \min(\exp(B'_p), 1).$$

Para estimar a medida de endocruzamento f , Ayres e Balding (1998) propuseram o algoritmo a seguir, que é a segunda parte do ciclo do MCMC.

4.2 Algoritmo para estimar a medida de endocruzamento

Bloco II:

1. Gere um candidato f' da distribuição geradora de candidatos h , que é uma *uniforme* com parâmetros

$$\max\left(-\frac{p'_{min}}{1-p'_{min}}, f - \varepsilon_f\right) \text{ e } \min(f + \varepsilon_f, 1),$$

onde o erro da medida de endocruzamento ε_f deve ser maior que $\frac{k^2 \varepsilon_p}{(k-1)(k-1-k\varepsilon_p)}$,

sendo ε_f escolhido de modo a se obter a **taxa de aceitação** (ou seja, contabiliza o número de vezes em que f' foi aceito, pelo algoritmo Metropolis, ao longo dos ciclos MCMC) entre 20% e 50% (Gilks; Richardson e Spiegelhater, 1996); k é o número de alelos.

- 1.1 Verifique se $f' \in \left[-\frac{p'_{min}}{1-p'_{min}}, 1\right]$. Em caso afirmativo, vá para (2), caso contrário vá para (1).

2. Defina a probabilidade de aceitação $\alpha(f, f')$ como

$$\alpha(f, f') = \min\left(\frac{t(f')h(f|f')}{t(f)h(f'|f)}, 1\right),$$

onde $t(f)$ e $t(f')$ representam a distribuição condicional completa *a posteriori* de f e de f' .

3. Gere um valor a de uma *uniforme* $(0, 1)$. Aceite f' se $a \leq \alpha(f, f')$.

Para o caso da medida de endocruzamento, considere a distribuição *a priori* de f dada por

$$\pi(f) = \begin{cases} \frac{1}{1 + \frac{p_{min}}{1 - p_{min}}} & , \text{ se } -\frac{p_{min}}{1 - p_{min}} \leq f \leq 1 \\ 0 & , \text{ caso contrário.} \end{cases}$$

A verossimilhança é a mesma da equação (4). A distribuição condicional completa *a posteriori* de f é, então, dada por

$$\begin{aligned} t(f) &= \pi(f|p_1, \dots, p_k; n_{11}, \dots, n_{kk}) \propto L(f, \{p_i\})\pi(f) \\ &= \prod_{i=1}^k \left[(fp_i + (1-f)p_i^2)^{n_{ii}} \prod_{j=i+1}^k (2p_i p_j (1-f))^{n_{ij}} \right] I_{\left[-\frac{p_{min}}{1-p_{min}}, 1\right]}(f) \quad (9) \end{aligned}$$

Como no Bloco I, é necessário calcular o termo do item (2), $B_f = \frac{t(f')h(f|f')}{t(f)h(f'|f)}$, para encontrar a probabilidade de aceitação. Assim, utilizando a equação (9), obtemos:

$$\begin{aligned} B_f &= \frac{\prod_{i=1}^k \left[(f'p_i + (1-f')(p_i)^2)^{n_{ii}} \prod_{j=i+1}^k (2p_i p_j (1-f'))^{n_{ij}} \right] \times I_{\left[-\frac{p_{min}}{1-p_{min}}, 1\right]}(f')}{\prod_{i=1}^k \left[(fp_i + (1-f)p_i^2)^{n_{ii}} \prod_{j=i+1}^k (2p_i p_j (1-f))^{n_{ij}} \right] \times I_{\left[-\frac{p_{min}}{1-p_{min}}, 1\right]}(f)} \\ &\quad \times \frac{I_{(LI', LS')}(f')}{I_{(LI, LS)}(f)} \quad , \quad (10) \end{aligned}$$

onde:

- $LI = \max\left(-\frac{p'_{min}}{1-p'_{min}}, f - \varepsilon_f\right)$ e $LI' = \max\left(-\frac{p'_{min}}{1-p'_{min}}, f' - \varepsilon_f\right)$, limites inferiores da distribuição geradora de candidatos h quando se considera f ou f' ;

- $LS = \min(f + \varepsilon_f, 1)$ e $LS' = \min(f' + \varepsilon_f, 1)$, limites superiores da distribuição geradora de candidatos h quando se considera f ou f' .

Supondo que $f \in \left[-\frac{p'_{min}}{1-p'_{min}}, 1\right] \cap (LI, LS)$ e $f' \in \left[-\frac{p'_{min}}{1-p'_{min}}, 1\right] \cap (LI', LS')$, é possível fazer uma transformação logarítmica em B_f , denominada por B'_f . Logo,

$$\begin{aligned} B'_f &= \sum_{i=1}^k n_{ii} \ln(f'p_i + (1-f')(p_i)^2) + \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_{ij} \ln(2p_i p_j (1-f')) \\ &\quad - \sum_{i=1}^k n_{ii} \ln(fp_i + (1-f)p_i^2) - \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_{ij} \ln(2p_i p_j (1-f)) \quad . \quad (11) \end{aligned}$$

É importante ressaltar que, como no caso de B'_p , cada termo $\ln(\cdot)$ na fórmula de B'_f é uma proporção genotípica. Dessa forma, deve-se considerar esta restrição no problema. Portanto, a probabilidade de aceitação mostrada no item (2) será dada por

$$\alpha(f, f') = \min(\exp(B'_f), 1).$$

5 Dados simulados a partir do modelo endogâmico

Neste trabalho utilizou-se dados simulados a partir do modelo endogâmico (vide expressão (1)) sob vários cenários, que diferiram:

- pelo tamanho da amostra de indivíduos selecionados na população, $n = 50$, 200 ou 1.000;
- pela intensidade do endocruzamento na população, traduzida por f :

$$f = -\frac{p_{min}}{1 - p_{min}} + 0.05, \quad f = 0.05 \text{ ou } f = 0.5;$$

- Pelo número de alelos presentes no *loco* em estudo, $k=2$, 6 ou 15.

Dessa forma, temos um total de 27 casos ou cenários a considerar.

5.1 Algoritmo para gerar as frequências genotípicas

Na geração de dados de frequências genotípicas, segundo o modelo endogâmico, que descreve um modelo multinomial, utilizamos o algoritmo dado a seguir. Com os quatro primeiros itens, definimos os parâmetros populacionais.

1. Gere k observações, o_1, \dots, o_k , a partir de uma distribuição *uniforme* $(0, 1)$, e some estes valores, obtendo um valor s .
2. Descreva as k proporções alélicas populacionais p_1, \dots, p_k , através de $p_i = o_i/s$, $i = 1, \dots, k$. Dessa forma, garante-se que $\sum_{i=1}^k p_i = 1$.
3. Fixe uma medida de endocruzamento, f^* , como se fosse a medida de endocruzamento verdadeira da população.
4. Calcule as proporções genotípicas a partir das proporções alélicas populacionais, p_i encontradas no item (2), e da medida de endocruzamento fixada no item anterior, f^* . Calcule as proporções genotípicas homocigotas e heterocigotas (vide expressão (1))
5. Gere as frequências genotípicas a partir de uma distribuição *Multinomial*, $MULT(n; p = (p_{11}, \dots, p_{kk}))$, sendo n o número de indivíduos amostrados, e p_{ii} e p_{ij} as proporções genotípicas homocigotas e heterocigotas encontradas no item (4).

Com base em dados simulados a partir deste algoritmo, estimou-se f pelos procedimentos Bayesianos e MCMC (Metropolis-Hastings), e pelos métodos clássicos de Nei e Chesser (NC) (1983) Robertson e Hill (RH) (1984). Nos casos de NC e RH, os intervalos de confiança foram obtidos com base em um procedimento Bootstrap não-paramétrico (vide Efron, 1979; Efron e Tibshirani, 1993) com amostras de tamanho 500.

5.2 Diagnósticos de convergência MCMC

No procedimento Bayesiano, utilizou-se 4.000 iterações com *burn-in* de 100 observações. Este número pequeno se deve ao fato do algoritmo convergir rapidamente. Um dos motivos é a restrição dos parâmetros envolvidos através dos seus limites. Após a obtenção das 4.000 observações, efetuou-se uma amostragem sistemática considerando saltos de tamanho 4. Assim, o tamanho amostral resultante foi de 975 observações. A partir dessa amostra foram feitos os testes de diagnóstico de Geweke (1992), Gelman e Rubin (1992) e de autocorrelação utilizando-se o programa CODA (<http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/>). Além disso, procedeu-se ao teste de Heidelberg e Welch (1983) para verificar a estacionariedade da cadeia e ao teste de Raftery-Lewis (1992) para se avaliar o número de observações necessárias no período de *burn-in*.

Um procedimento que também é útil na avaliação da convergência da cadeia de Markov para uma distribuição de equilíbrio (possivelmente a distribuição *a posteriori*) é a construção de um gráfico da densidade *a posteriori* utilizando núcleo estimador (Kernel). A Figura 1 ilustra uma saída do CODA, onde a densidade *a posteriori* é aproximada por núcleo estimador. Observa-se que há uma boa convergência dos valores gerados de f . Maiores detalhes sobre núcleo estimador podem ser vistos em Silverman (1986).

Os testes de diagnósticos mostraram que, em todos os casos, a cadeia convergiu e apresenta características de uma cadeia estacionária. Segundo o procedimento de Raftery-Lewis (1992), o maior período de *burn-in* foi 24. Isto mostra que o *burn-in* estipulado, de 100 observações, está mais do que adequado, podendo, até, ser considerado conservador.

A Figura 2 descreve um dos gráficos da distribuição *a posteriori* da medida de endocruzamento para o método de Ayres e Balding (1998) considerando o caso de dois alelos, $f^* = -0.95$ e 50 indivíduos. Observe que a distribuição *a posteriori* é simétrica. Os demais gráficos foram omitidos por serem semelhantes à Figura 2.

5.3 Análise dos dados simulados

Na avaliação das propriedades amostrais dos estimadores de f , apresentados nas seções anteriores, para cada um dos 27 casos descritos no início da Seção 5, simulamos m amostras e, para cada uma, estimamos f por meio de estimativas pontuais (média *a posteriori*) e por intervalos. Por demandar um tempo

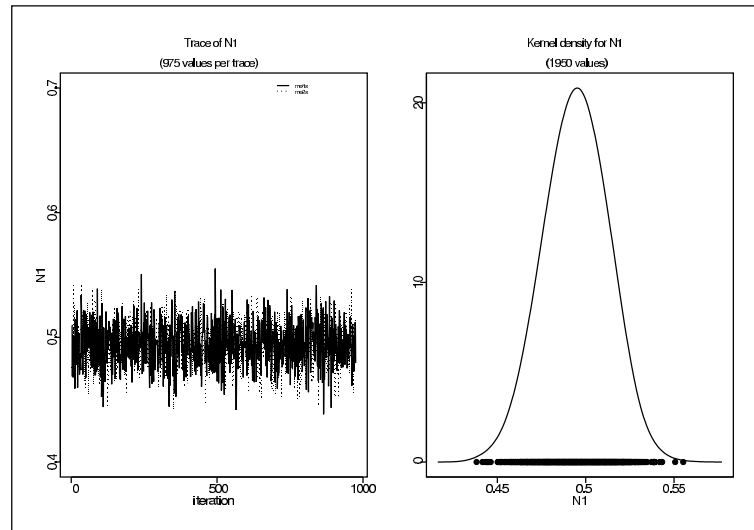


Figura 1 - Gráfico de linhas e da densidade de Kernel para o caso de 15 alelos, $f^* = 0.5$ e 1.000 indivíduos.

computacional muito elevado, só foi possível simular e analisar $m = 20$ amostras para cada um dos 27 casos. Para se ter um exemplo, apenas no caso de $k = 15$, $n = 1.000$ indivíduos e $f = 0.5$, o tempo computacional total despendido com a geração dos dados, estimação Bayesiana, estimação freqüentista e o procedimento Bootstrap para $B = 500$ amostras, ficou em torno de 20 horas.

De modo a avaliar o desempenho dos estimadores de f , análises comparativas de medidas-resumo que estimam locação e dispersão dos estimadores de f , foram empreendidas. Além disso, de modo a verificar a qualidade dos estimadores por intervalo, estimamos a probabilidade de cobertura.

Portanto, para cada um dos 27 casos, estimamos o Vício, o Erro Quadrático Médio (EQM) e a Probabilidade estimada de cobertura dos intervalos de confiança e de credibilidade de f . Os resultados dessas análises proporcionam alguns indicativos da eficiência de cada método analisado, e estão sumarizados nas Tabelas 1 a 3.

Valores estimados de f : Nitidamente os valores estimados da medida de endocruzamento, f , que estão mais próximos da medida de endocruzamento verdadeira, f^* , na maioria dos casos foram as de Robertson e Hill (RH) (1984) e de Nei e Chesser (NC) (1983), sendo a de RH ligeiramente superior à de NC. O Método de Ayres e Balding (AB) (1998) superestima o valor de f^* quando este está muito próximo do limite inferior do espaço paramétrico, possivelmente por considerar a restrição dos limites de f .

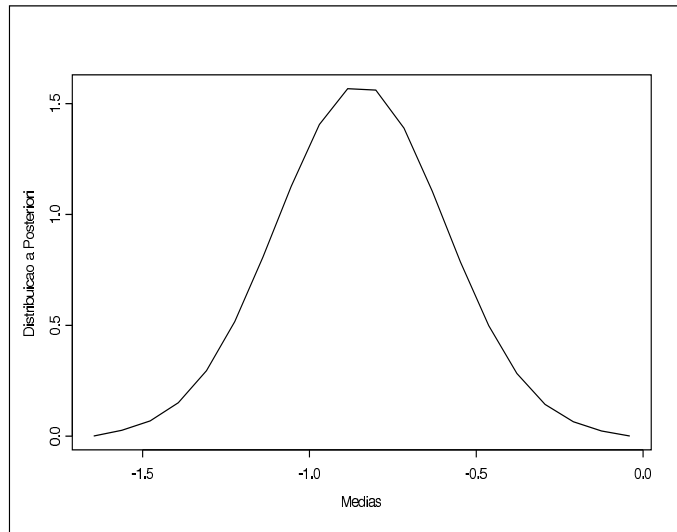


Figura 2 - Gráfico da distribuição *a posteriori* da medida de endocruzamento calculada pelo método de Ayres e Balding considerando-se o caso de dois alelos, $f^* = -0.95$ e 50 indivíduos.

Vício: Os vícios estimados foram os menores, na maioria dos casos, para o método de Robertson e Hill (RH) (1984), seguido pelo método de Nei e Chesser (NC) (1983), que muitas vezes forneceu resultados muito próximos daqueles observados para o método de RH. Já os vícios estimados de \hat{f} obtidos a partir do método de Ayres e Balding (AB) (1998), foram, na maioria dos casos, maiores do que os proporcionados pelos outros métodos. Em alguns casos, estes vícios são muito maiores do que os observados nos métodos anteriores.

Erro Quadrático Médio (EQM): Para f^* próximo do limite inferior do espaço paramétrico, os menores valores de EQM foram observados para o método de Robertson e Hill (RH) (1984). Para f^* positivo, os menores valores de EQM foram observados para o método de Ayres e Balding (1998).

Probabilidade estimada de cobertura: Para o caso de $k = 2$ alelos e $n = 50$ indivíduos, a probabilidade estimada de cobertura não atinge o nível nominal para nenhum dos métodos utilizados. Porém, o método de Ayres e Balding (AB) (1998) proporciona a probabilidade estimada de cobertura mais próxima do nível nominal de 95%. Para $n = 200$ e $n = 1.000$, os três métodos apresentam valores estimados de probabilidade de cobertura muito semelhantes.

Para f^* próximo do limite inferior do espaço paramétrico de f e para $k = 6$ e $k = 15$ alelos, independentemente do tamanho da amostra n , o método de Ayres e Balding (AB) (1998) proporcionou os menores valores estimados de probabilidade de cobertura. Para f^* positivo, independentemente do tamanho da amostra n , todos os métodos proporcionam probabilidade de cobertura estimada muito semelhantes e próximas ao nível nominal 95%. Também, como era de se esperar, à medida que o número de indivíduos aumenta, a probabilidade estimada de cobertura se aproxima de 1.

Para se ter uma idéia mais clara de quão comparáveis são as inferências sobre f proporcionadas por cada método, foram feitos gráficos de dispersão dos valores estimados de f para cada um dos casos analisados. O índice nominal adotado na construção dos intervalos de confiança e de credibilidade foi de 95%. Em cada gráfico (ver, por exemplo, a Figura 3) traçamos no eixo horizontal uma linha que descreve o valor verdadeiro (populacional) de f (fixado, nas simulações, em $f = -0.95$) e ao redor da qual os valores estimados de f estão dispostos. Por economia de espaço mostramos apenas um desses gráficos.

Com base nesses gráficos observou-se que, para $n = 50$ e f^* próximo ao limite inferior do espaço paramétrico de f , independente do número k de alelos, o método de Ayres e Balding (AB) (1998) proporcionou valores estimados de f muito discrepantes do valor verdadeiro, enquanto os valores estimados de f obtidos pelos métodos de Robertson e Hill (RH) (1984) e de Nei e Chesser (NC) (1983) proporcionaram valores estimados de f muito mais próximos do valor verdadeiro, sendo estes valores estimados muito semelhantes. Observou-se também que à medida que n cresce, as diferenças nos valores estimados de f , proporcionadas pelos diferentes métodos, tornam-se cada vez menores. Porém, o método de Ayres e Balding (AB) (1998) persistentemente acarreta vícios positivos na estimação de f .

Observamos também que independente do número k de alelos, para $f^* = 0.05$, que representa um pequeno desvio do equilíbrio de Hardy-Weinberg, não se observa superioridade de nenhum dos métodos. A medida que o número k de alelos aumenta, as discrepâncias entre os valores estimados de f e o valor verdadeiro, f^* , diminuem. Além disso, independente do número k de alelos, para $f^* = 0.50$, isto é, para uma população com alta prevalência de cruzamentos endogâmicos, não se observou nenhuma superioridade de nenhum dos métodos. Um caso atípico foi o de $f^* = 0.50$ e $k = 15$ alelos, no qual o método de Robertson e Hill (RH) subestima f^* .

Conclusões e propostas futuras

Em genética de populações, a Lei de Hardy-Weinberg é de fundamental importância no estudo de genética de populações. Uma forma de se medir o desequilíbrio de Hardy-Weinberg é pela medida de endocruzamento. Há vários métodos na literatura para a estimação da medida de endocruzamento.

Neste trabalho, para a maioria dos casos em que simulamos dados sob o modelo endogâmico, os métodos de Robertson e Hill (1984) e de Nei e Chesser (1983) apresentaram valores estimados próximos da medida de endocruzamento verdadeira, sendo o método de Robertson e Hill (1984) ligeiramente superior ao de Nei e Chesser (1983). Por outro lado, os valores estimados de f obtidos por esses dois métodos podem ultrapassar consideravelmente os limites teóricos de f . Além disso, tais medidas não consideram, de forma explícita, o modelo endogâmico.

Devido a tais limitações, Ayres e Balding (1998) propuseram um método alternativo na tentativa de sanar tais deficiências. No entanto, como descrito na Seção 5, na qual fazemos uma análise comparativa dos três métodos baseada em dados simulados segundo o modelo endogâmico, observou-se que o método de Ayres e Balding (1998) é, de fato, eficiente no que diz respeito à produção de estimativas de f que respeitem os limites teóricos do parâmetro. Porém, observou-se que o método de Ayres e Balding (1998) pode produzir valores estimados de f com vícios apreciáveis, sendo de magnitude muito superior àqueles produzidos pelos demais métodos. No entanto, para casos onde a medida de endocruzamento é positiva, o método de Ayres e Balding (1998) apresenta boa *performance*.

Neste trabalho, nas análises que foram feitas para se estimar as medidas de endocruzamento, considerou-se que os *locos* são independentes, isto é, considerou-se a ausência de influências conjuntas dos *locos* no cálculo da medida de endocruzamento f . Tal pressuposto não é realístico para a maioria das situações práticas. Um modelo que considera tais influências pode revelar aspectos importantes na análise de dados reais. O modelo multiloco, proposto por Ayres e Balding (2001), pode ser um ponto de partida no caso Bayesiano. No caso freqüentista, pode-se utilizar as medidas de endocruzamento multilocos proposta por Nei e Chesser (1983).

Agradecimentos

As autoras agradecem aos pareceristas por suas críticas e sugestões que muito ajudaram a aperfeiçoar o artigo.

da-SILVA, C. Q.; ARMBORST, T. Methods for measuring departures from the Hardy-Weinberg equilibrium through endogamy measures. *Rev. Mat. Estat.*, São Paulo, v.24, n.4, p.51-71, 2006.

- **ABSTRACT:** One of the main measures of departure from the Hardy-Weinberg Equilibrium (HWE) is the endogamy coefficient, f . Such quantity measures heterozygotes deficit in the population. In this study three methods for estimating f are used: the Ayres and Balding's (AB) (1998) Bayesian method, and two classic ones, by Nei and Chesser (NC)(1983) and by Robertson and Hill(RH)(1984). We analyzed genotype frequency simulated data considering samples sizes 50, 200 and 1,000 and 2, 6 and 15 alleles, with $n = 50$ being the most realistic case. The methods were evaluated through the comparison of credibility intervals (via MCMC) and confidence intervals (calculated via Bootstrap). We found that the Ayres and Balding method is very efficient in producing estimates of f that lie in the parameter space. Moreover, the AB method has the advantage of allowing the calculation of allele frequencies. However, not only does the AB method demand too complex calculations, but can also produce very biased estimates of f , which are much larger than the ones observed for the other two methods.
- **KEYWORDS:** Genetics; endogamy measures; MCMC; classic and Bayesian inference.

Referências

- 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.
- AYRES, K. L.; BALDING, D. J. Measuring gametic gesequilibrium from multilocus data. *Genetics*, Bethesda, v.157, p.413-423, 2001.
- BERNSTEIN, F. Zusammenfassende Betrachtungen über die erblichen Blüten-structuren des Menschen. *Z. indukt. Abstamm.-u. VererbLehre*, v.37, p.237-270, 1925.
- CONVERGENCE diagnostics and output analysis for bugs output. Plummer, M. *MRC biostatistics unit*. versão 0.40, 1995. Disponível em: <<http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml>>. Acesso em 2005.
- EFRON, B. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, Hayward, v.7, p.1-26, 1979.
- EFRON, B.; TIBSHIRANI, R. J. *An introduction to the bootstrap*: monograph on statistics and applied probability. New York: Chapman & Hall, 1993. p.57.
- GARDNER, E. J. *Genética*. 5. ed. Utah: Interamericana, Utah State University. 1977. 515p.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, Hayward, v.7, p.457-72, 1992.
- GELMAN, A. et al. *Bayesian data analysis. Tests in statistical science series*. 2003.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Stat.*, Oxford, v.4, p.169-194, 1992.

- GILKS, W. R.; RICHARDSON, S.; SPIGELHATER, D. J. *Markov chain Monte Carlo in practice*. New York: Chapman & Hall, 1996. p.1-9.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, London, v.57, p.97-109, 1970.
- HEIDELBERG, P.; WELCH, P. Simulation run length control in the presence of an initial transient. *Oper. Res.*, Maltimore, v.31, p.1109-1144, 1983.
- METROPOLIS, N. et al. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, Woodbury, v.21, p.1087-1092, 1953.
- MURRAY, B. W. The estimation of genetic distance and population substructure from microsatellite allele frequency data. McMaster University, 1996. Disponível em: <helix.biology.mcmaster.ca/brent/brent.html>. Acesso em 1996.
- NEI, M.; CHESSER, R. K. Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.*, Cambridge, v.47, p.253-259, 1983.
- RAFTERY, A. L.; LEWIS, S. How many iterations in the Gibbs sampler? *Bayesian Stat.*, Oxford, v.4, p.763-74, 1992. 515p.
- ROBERTSON, A.; HILL, W. G. Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics*, Bethesda, v.107, p.703-718, 1984.
- SILVERMAN, B. W. *Density estimation for statistics and data analysis*. New York: Chapman and Hall, 1986. 175p.

Recebido em 22.02.2005.

Aprovado após revisão em 10.12.2006.

Tabela 1 - Avaliação da acurácia do estimador pela estimativa, vício, EQM e probabilidade de cobertura (índice nominal de 95%) para o caso de 2 alelos

f^*	Número de indivíduos	Avaliação desempenho	Métodos		
			\hat{f}_{ab}	\hat{f}_{nc}	\hat{f}_{rh}
-0.95	50	Estimativa	-0.84630	-0.94623	-0.94673
		Vício	0.10370	0.00377	0.00327
		EQM	0.01183	0.00132	0.00129
		Prob. Cobertura	0.85	0.80	0.80
	200	Estimativa	-0.90412	-0.94710	-0.94722
		Vício	0.04588	0.00290	0.00278
		EQM	0.00238	0.00032	0.00031
		Prob. Cobertura	0.95	0.95	0.95
	1.000	Estimativa	-0.92257	-0.94968	-0.94971
		Vício	0.02743	0.00032	0.00029
		EQM	0.00079	0.00004	0.00004
		Prob. Cobertura	1.00	1.00	1.00
0.05	50	Estimativa	0.05949	0.01838	0.01876
		Vício	0.00949	-0.03162	-0.03124
		EQM	0.01940	0.02034	0.02076
		Prob. Cobertura	0.95	0.80	0.80
	200	Estimativa	0.05343	0.04899	0.04914
		Vício	0.00343	-0.00101	-0.00086
		EQM	0.00654	0.00680	0.00683
		Prob. Cobertura	0.95	0.95	0.95
	1.000	Estimativa	0.04401	0.04405	0.04408
		Vício	-0.00599	-0.00595	-0.00592
		EQM	0.00115	0.00119	0.00119
		Prob. Cobertura	0.95	0.90	0.90
0.5	50	Estimativa	0.47282	0.49596	0.48089
		Vício	-0.02718	-0.00404	-0.01911
		EQM	0.00929	0.00965	0.01708
		Prob. Cobertura	0.95	0.90	0.90
	200	Estimativa	0.48917	0.50402	0.49938
		Vício	-0.01083	0.00402	-0.00062
		EQM	0.00137	0.00137	0.00355
		Prob. Cobertura	1.00	1.00	1.00
	1.000	Estimativa	0.49134	0.50459	0.50497
		Vício	-0.00866	0.00459	0.00497
		EQM	0.00067	0.00069	0.00069
		Prob. Cobertura	1.00	1.00	1.00

Denota-se: \hat{f}_{AB} : \hat{f} pelo método Bayesiano de Ayres e Balding (1998);

\hat{f}_{NC} : \hat{f} pelo método clássico de Nei e Chesser (1983);

\hat{f}_{RH} : \hat{f} pelo método clássico de Robertson e Hill (1984).

Tabela 2 - Avaliação da acurácia do estimador pela estimativa, vício, EQM e probabilidade de cobertura (índice nominal de 95%) para o caso de 6 alelos

f^*	Número de indivíduos	Avaliação desempenho	Métodos		
			\hat{f}_{ab}	\hat{f}_{nc}	\hat{f}_{rh}
-0.15	50	Estimativa	-0.08252	-0.15690	-0.14698
		Vício	0.06748	-0.00690	0.00302
		EQM	0.00496	0.00117	0.00090
		Prob. Cobertura	0.35	0.85	0.85
	200	Estimativa	-0.12490	-0.14762	-0.14614
		Vício	0.02510	0.00238	0.00386
		EQM	0.00081	0.00031	0.00026
		Prob. Cobertura	0.75	1.00	1.00
	1.000	Estimativa	-0.13865	-0.14615	-0.14604
		Vício	0.01135	0.00385	0.00396
		EQM	0.00015	0.00007	0.00006
		Prob. Cobertura	0.75	0.95	0.95
0.05	50	Estimativa	0.10020	0.06348	0.07277
		Vício	0.05020	0.01348	0.02277
		EQM	0.00583	0.00574	0.00620
		Prob. Cobertura	1.00	1.00	0.90
	200	Estimativa	0.06935	0.06408	0.05820
		Vício	0.01935	0.01408	0.00820
		EQM	0.00182	0.00218	0.00190
		Prob. Cobertura	1.00	0.85	0.85
	1.000	Estimativa	0.05203	0.05198	0.04853
		Vício	0.00203	0.00198	-0.00147
		EQM	0.00031	0.00036	0.00029
		Prob. Cobertura	1.00	0.85	0.95
0.5	50	Estimativa	0.48785	0.50700	0.49193
		Vício	-0.01215	0.00700	-0.00807
		EQM	0.00860	0.00981	0.01360
		Prob. Cobertura	0.95	0.90	0.95
	200	Estimativa	0.47840	0.49202	0.48394
		Vício	-0.02160	-0.00798	-0.01606
		EQM	0.00192	0.00167	0.00249
		Prob. Cobertura	1.00	1.00	0.90
	1.000	Estimativa	0.48916	0.50162	0.49987
		Vício	-0.01084	0.00162	-0.00013
		EQM	0.00052	0.00043	0.00054
		Prob. Cobertura	1.00	0.95	1.00

Denota-se: \hat{f}_{AB} : \hat{f} pelo método Bayesiano de Ayres e Balding (1998);

\hat{f}_{NC} : \hat{f} pelo método clássico de Nei e Chesser (1983);

\hat{f}_{RH} : \hat{f} pelo método clássico de Robertson e Hill (1984).

Tabela 3 - Avaliação da acurácia do estimador pela estimativa, vício, EQM e probabilidade de cobertura (índice nominal de 95%) para o caso de 15 alelos

f^*	Número de indivíduos	Avaliação desempenho	Métodos		
			\hat{f}_{ab}	\hat{f}_{nc}	\hat{f}_{rh}
0.05	50	Estimativa	0.02116	-0.03212	-0.02798
		Vício	0.04259	-0.01069	-0.00655
		EQM	0.00206	0.00116	0.00058
		Prob. Cobertura	0.60000	0.90000	0.80000
	200	Estimativa	-0.00275	-0.02322	-0.02266
		Vício	0.01868	-0.00179	-0.00123
		EQM	0.00042	0.00032	0.00023
		Prob. Cobertura	0.75000	0.90000	0.90000
	1.000	Estimativa	-0.01692	-0.02287	-0.02253
		Vício	0.00451	-0.00145	-0.00110
		EQM	0.00004	0.00005	0.00004
		Prob. Cobertura	0.95000	0.95000	0.95000
0.5	50	Estimativa	0.07043	0.02642	0.03674
		Vício	0.02043	-0.02358	-0.01326
		EQM	0.00239	0.00411	0.00505
		Prob. Cobertura	1.00000	0.90000	0.85000
	200	Estimativa	0.05881	0.05420	0.04719
		Vício	0.00881	0.00420	-0.00281
		EQM	0.00076	0.00088	0.00089
		Prob. Cobertura	1.00000	0.85000	0.85000
	1.000	Estimativa	0.05287	0.05262	0.05005
		Vício	0.00287	0.00262	0.00005
		EQM	0.00017	0.00019	0.00021
		Prob. Cobertura	1.00000	1.00000	0.85000
0.95	50	Estimativa	0.47033	0.49017	0.43544
		Vício	-0.02967	-0.00983	-0.06456
		EQM	0.00490	0.00443	0.00852
		Prob. Cobertura	0.90000	0.90000	0.85000
	200	Estimativa	0.48579	0.49913	0.47937
		Vício	-0.01421	-0.00087	-0.02063
		EQM	0.00131	0.00117	0.00370
		Prob. Cobertura	1.00000	1.00000	0.85000
	1.000	Estimativa	0.49276	0.50511	0.50211
		Vício	-0.00724	0.00511	0.00211
		EQM	0.00028	0.00027	0.00080
		Prob. Cobertura	1.00000	0.95000	0.90000

Denota-se: \hat{f}_{AB} : \hat{f} pelo método Bayesiano de Ayres e Balding (1998);

\hat{f}_{NC} : \hat{f} pelo método clássico de Nei e Chesser (1983);

\hat{f}_{RH} : \hat{f} pelo método clássico de Robertson e Hill (1984).

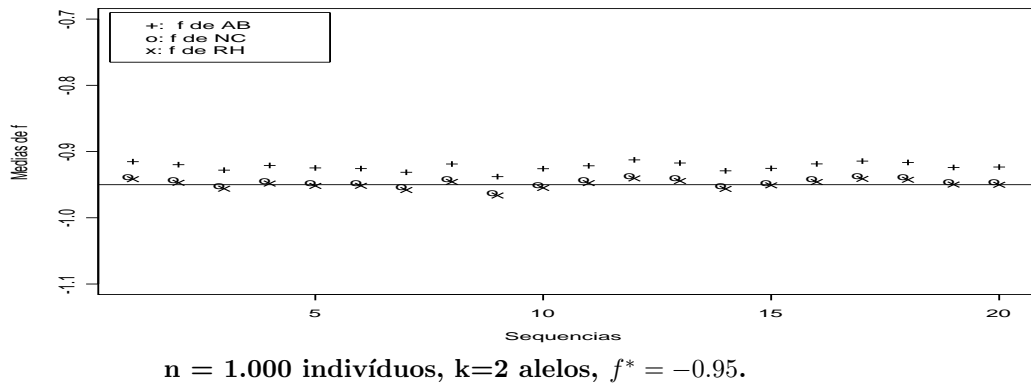
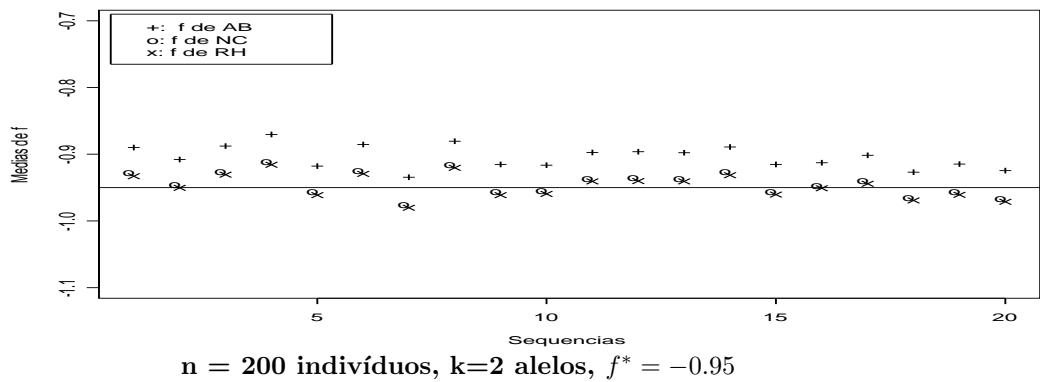
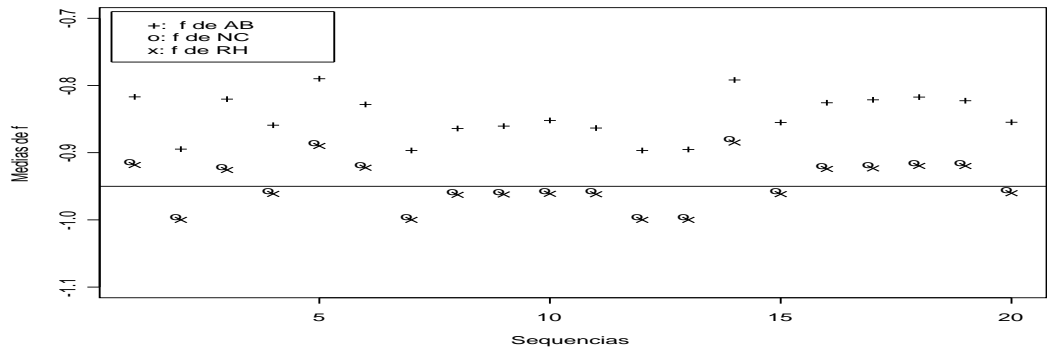


Figura 3 - Gráficos de dispersão para o caso de $k = 2$ alelos e medida de endocruzamento verdadeira $f^* = -0.95$. Notação: “+ f de AB”: \hat{f} pelo Método de Ayres e Balding (1998); “o f de NC”: \hat{f} pelo Método de Nei e Chesser (1983); “x f de RH”: \hat{f} pelo Método de Robertson e Hill (1984).