



Predição de valores genéticos sob heterogeneidade de variâncias utilizando inferência Bayesiana e clássica

José Marques Carneiro Júnior¹, Giselle Mariano Lessa de Assis², Jonathas Vasconcelos de Melo³, Eriane da Silva Nobre⁴, Ricardo Frederico Euclides⁵

¹Pesquisador da Embrapa Acre. e-mail: marques@cpafac.embrapa.br

²Pesquisadora da Embrapa Acre. e-mail: giselle@cpafac.embrapa.br

³Graduando do curso de Engenharia Agrônoma da Universidade Federal do Acre. e-mail: melo.joe@gmail.com

⁴Graduanda do curso de Engenharia Agrônoma da Universidade Federal do Acre. e-mail: enaire_20@hotmail.com

⁵Professor do Departamento de Zootecnia - UFV. Viçosa-MG. e-mail: rbaja@ufv.br

Resumo: A maioria das pesquisas demonstra que a heterogeneidade de variâncias é um problema na estrutura dos dados, resultando em classificação incorreta dos animais, devido à baixa acurácia da predição dos valores genéticos. Assim, verificou-se o comportamento das metodologias EBLUP e Bayesiana, em populações com diferentes tamanhos na presença de homogeneidade e heterogeneidade de variâncias. Para cada estrutura foram simulados dois tamanhos de população. Para a metodologia Bayesiana foram utilizados três níveis de informação: *priors* não-informativos, *priors* pouco informativos e *priors* informativos. A acurácia da predição foi obtida por meio do Quadrado Médio do Erro e da Correlação de Spearman. Foi observado que as populações de tamanho grande com homogeneidade de variâncias, de modo geral, conduziram a menores valores de Quadrado Médio do Erro e maiores Correlações de Spearman, independente da metodologia empregada. Os maiores Quadrado Médios dos Erros e as menores Correlações de Spearman foram observados para as populações pequenas com presença de heterogeneidade. Nesta situação, as correlações de Spearman foram menores que 0,40, demonstrando alterações na classificação dos indivíduos. Concluiu-se que a predição dos valores genéticos dos animais é influenciada negativamente pela presença de heterogeneidade de variância genética e ambiental e pelo tamanho reduzido da população. Nestas condições, as metodologias EBLUP e Bayesiana conduzem a predições similares de valores genéticos.

Palavras-chave: Bayesiana, componentes de variância, heterogeneidade de variâncias, informação *a priori*, simulação, valores genéticos

Breeding values prediction under heterogeneity of variance using Bayesian and classical inference

Abstract: Most of the studies show that heterogeneity of variance is a data structure problem, resulting in incorrect classification of animals, due the low accuracy of breeding values prediction. Thus, the comportment of EBLUP and Bayesian methodology was verified, in populations with different size in the presence of homogeneity and heterogeneity variance. Two population sizes were simulated for each variance structure. Three information levels were used for Bayesian methodology: no-informative priors; slightly informative priors and informative priors. The prediction accuracy was obtained by Mean Square Error and Spearman Correlation. It was observed that small size populations with homogeneity of variance, in general, conduced to smaller Mean Square Error and largest Spearman Correlation, independent of the methodology employed. The largest Mean Square Error and the smaller Spearman Correlation were observed to small populations with heterogeneity. In this situation, the Spearman correlations were smaller that 0.40, demonstrating alterations in the individuals classification. It was concluded that breeding values prediction is influenced negatively by genetic and environmental heterogeneity of variance and by reduced population size. In this condition, EBLUP and Bayesian methodologies lead a similar breeding values prediction.

Keywords: Bayesian; breeding value, heterogeneity of variance, *priori* information, variance components, simulation

Introdução

A maioria das pesquisas demonstra ser a heterogeneidade um problema na própria estrutura dos dados, que pode resultar em uma classificação errônea dos animais, diminuindo o progresso genético. Quando a heterogeneidade não é considerada, diferenças de variâncias dentro das subclasses podem resultar na predição de valores genéticos viesados, redução no progresso genético e desproporcional número de animais selecionados de ambientes com diferentes variâncias (Weigel e Gianolla, 1992).

A Metodologia de Modelos Mistos proposta por Henderson (1973) vem sendo amplamente empregada no melhoramento animal para obtenção do Melhor Preditor Linear Não-viesado (BLUP) dos valores genéticos. Porém, essa metodologia pressupõe que os componentes de variância sejam conhecidos, o que raramente ocorre na prática. A estratégia de análise utilizada quando tais componentes são desconhecidos consiste na substituição dos valores paramétricos dos componentes de variância por suas estimativas de Máxima Verossimilhança Restrita (REML), recebendo a denominação de *BLUP Empírico* (EBLUP).

Métodos Bayesianos vêm sendo utilizados como uma opção para a solução de problemas relacionados à avaliação de mérito genético em populações animais. O *Gibbs Sampling* é aplicável à estimação de componentes de variância e predição dos valores genéticos, permitindo, por suas propriedades, a inferência Bayesiana.

Dentro do contexto apresentado, os objetivos desse trabalho foram:

- Verificar o efeito da heterogeneidade de variâncias e do tamanho da população na predição dos valores genéticos;
- Verificar a influência do nível de informação *a priori* ao utilizar metodologia Bayesiana, para predição dos valores genéticos; e
- Comparar a metodologia clássica EBLUP com a Bayesiana na análise genética em dados com presença heterogeneidade de variâncias.

Material e Métodos

Os dados utilizados neste trabalho foram simulados por meio do sistema computacional GENESYS (Euclides, 1996). Na Tabela 1 são apresentados os códigos utilizados para representar as populações com diferentes tipos de estruturas e tamanhos de população simulados.

Tabela 1 Código utilizado para as populações simuladas com diferentes estruturas de variâncias e tamanhos.

Estrutura dos dados	Nível de Variabilidade	Tamanho da população		Nº de Indivíduos	
		População Pequena (P)	População Grande (G)	(P)	(G)
Homogeneidade de Variâncias (HO)	População 1 (1)	HO-P-1	HO-G-1	100	1.600
	População 2 (2)	HO-P-2	HO-G-2	100	1.600
	População 3 (3)	HO-P-3	HO-G-3	100	1.600
	Total (T)	HO-P-T	HO-G-T	300	4.800
Heterogeneidade de variância Genética e Ambiental (HGA)	Alta (A)	HGA-P-A	HGA-G-A	100	1.600
	Média (M)	HGA-P-M	HGA-G-M	100	1.600
	Baixa (B)	HGA-P-B	HGA-G-B	100	1.600
	Todas (T)	HGA-P-T	HGA-G-T	300	4.800

$T = (A+M+B)$ ou $T = (1 + 2 + 3)$

Os valores genéticos foram preditos pela metodologia EBLUP e Bayesiana. Nesta última foram considerados três diferentes níveis de informação *a priori*: *priors* informativos; *priors* pouco informativos e *priors* não-informativos. O seguinte modelo linear misto unicaracterístico foi empregado nas análises:

$$Y_{ij} = \mu + ef_i + a_{ij} + e_{ij},$$

em que y_{ij} é a observação fenotípica do j -ésimo animal no i -ésimo rebanho; μ é uma constante inerente a todas as observações; ef_i é o efeito do i -ésimo rebanho; a_{ij} é o efeito aleatório genético aditivo do j -ésimo animal no i -ésimo rebanho; e_{ij} é o efeito aleatório ambiental. Os valores genéticos preditos foram comparados com os valores genéticos reais, por meio da Correlação de Spearman e do Quadrado Médio do Erro, como em Harvey (1979), citado por Schenkel (1998).

$$QME = \sum_{i=1}^n \frac{1}{n} (\hat{a}_i - a_i)^2$$

em que QME é o quadrado médio do erro, n é o número de indivíduos, $\hat{a}_i$ refere-se ao valor genético predito e a_i ao valor genético real.

Resultados e Discussão

Pode ser observado na Tabela 2 que os menores Quadrados Médio dos Erros e as maiores Correlações de Spearman foram obtidos para as populações de tamanho grande com homogeneidade de variâncias, indicando que a presença de heterogeneidade de variâncias genética e ambiental reduz a acurácia da predição dos valores genéticos. Os resultados foram similares para as duas metodologias empregadas, sendo que a utilização de *priors* mais informativos pela metodologia Bayesiana, em todas as situações, não proporcionou melhorias na classificação dos indivíduos.

Tabela 2 QME e Correlação de Spearman (em parênteses) entre os valores genéticos reais e os preditos pela metodologia EBLUP e Bayesiana sob três níveis de informação, para as populações totais e subpopulações com três níveis de variabilidade e dois tamanhos nas diferentes estruturas de variância.

Populações	EBLUP	Bayesiana		
		NI	PI	I
Homogeneidade de variâncias				
HO-P-1	FC (FC)	167,63 (0.48)	126,82 (0.50)	118,11 (0.50)
HO-P-2	144,96 (0,60)	139,55 (0.60)	127,44 (0.59)	128,14 (0.59)
HO-P-3	100,79 (0,69)	168,84 (0.68)	100,96 (0.69)	101,28 (0.68)
HO-P-T	139,93 (0,59)	137,45 (0.61)	122,98 (0.64)	110,35 (0.66)
HO-G-1	126,51 (0,75)	125,27 (0.75)	125,45 (0.75)	124,59 (0.75)
HO-G-2	113,87 (0,78)	113,36 (0.78)	113,40 (0.78)	113,23 (0.78)
HO-G-3	128,44 (0.75)	129,58 (0.75)	127,41 (0.75)	125,42 (0.75)
HO-G-T	113,19 (0.78)	113,40 (0.78)	113,33 (0.78)	112,32 (0.78)
Heterogeneidade de Variâncias Genética e ambiental				
HGA-P-A	233,91(0.41)	222,70 (0.38)	214,54 (0.40)	213,67 (0.40)
HGA-P-M	FC(FC)	126,85 (0.40)	116,67 (0.41)	113,78 (0.41)
HGA-P-B	82,72 (0.36)	84,18 (0.34)	82,23 (0.35)	82,43 (0.35)
HGA-P-T	155,50 (0.42)	146,32 (0.43)	134,05 (0.43)	129,82 (0.44)
HGA-G-A	248,09 (0.52)	243,39 (0.52)	244,15 (0.52)	242,33 (0.52)
HGA-G-M	148,55 (0.49)	145,98 (0.49)	146,46 (0.49)	146,19 (0.49)
HGA-G-B	100,43 (0.45)	97,06 (0.46)	97,12 (0.46)	96,13 (0.46)
HGA-G-T	154,40 (0.54)	152,62 (0.54)	152,82 (0.54)	151,97 (0.54)

NI= Não informativo; PI = Pouco informativo; I = Informativo, FC = Falhou a convergência.

A metodologia Bayesiana obteve melhores resultados nas situações onde a metodologia REML não obteve convergência. O Quadrado Médio do Erro e a Correlação de Spearman obtidos pela Bayesiana foram similares aos das demais subpopulações do grupo estudado. Observa-se também que o nível de informação *a priori* reduziu levemente o Quadrado Médio do Erro, porém não alterou a classificação dos indivíduos.

Conclusões

A predição dos valores genéticos dos animais é influenciada negativamente pela presença de heterogeneidade de variância genética e ambiental e pelo tamanho reduzido da população.

Maiores níveis de informação conduzem a valores genéticos mais acurados, porém não alteram a classificação dos indivíduos.

As metodologias EBLUP e Bayesiana conduzem a resultados semelhantes de valores genéticos. Contudo, em conjuntos de dados pequenos a metodologia EBLUP pode não obter convergência.

Literatura citada

- EUCLYDES, R. F. **Uso do sistema para simulação Genesys na avaliação de métodos de seleção clássicos e associados a marcadores moleculares**. Viçosa: Universidade Federal de Viçosa - 1996. 150p. Tese (Doutorado em Genética e Melhoramento) _ Universidade Federal de Viçosa - 1996.
- SCHENKEL, F. S. **Studies on effects of parental selection on estimation of genetic parameters an breeding values of metric traits**. University of Ghelph, Ghelph, Canada. 1998. (Ph.d. Thesis).
- WEIGEL, D. A., GIANOLA, D., Estimation of Heterogeneous Within-Herd Variance Components Using Empirical Bayes Methods: A Simulation Study. **Journal of Dairy Science**, v. 75, p. 2824-2833, 1992.