



Estimação de componentes de variância utilizando-se inferência Bayesiana e frequentista em dados simulados sob heterogeneidade de variâncias

José Marques Carneiro Júnior¹, Giselle Mariano Lessa de Assis¹, Ricardo Frederico Euclides², Robledo de Almeida Torres², Paulo Sávio Lopes²

¹ Embrapa Acre, Rodovia BR 364, km 14, CP 321, 69908-770, Rio Branco, AC.

² Departamento de Zootecnia - Universidade Federal de Viçosa, CEP: 36571-000, Viçosa, MG.

RESUMO - Foi simulado um genoma de 3.000 centimorgans de comprimento considerando uma única característica quantitativa, governada por 800 locos com dois alelos por loco. Segundo a estrutura genômica proposta, foram simulados 1.500 machos e 1.500 fêmeas que formaram a população-base. A partir da população-base foram formadas duas populações iniciais, uma grande e outra pequena. Dois tipos de estruturas de heterogeneidade de variâncias foram inseridos nas populações iniciais: heterogeneidade de variância genética aditiva e heterogeneidade de variâncias genética aditiva e ambiental. Para obtenção destas estruturas, foram feitos descartes estratégicos dos valores genéticos aditivos e ambientais de acordo com o tipo de heterogeneidade e o nível de variabilidade desejada: alta, média ou baixa. Os componentes de variância foram estimados por meio da metodologia Bayesiana via Amostragem de Gibbs e pelo método REML. Para a metodologia Bayesiana, foram utilizados três níveis de informação *a priori*: não-informativo, pouco informativo e informativo. Os métodos comparados apresentaram resultados semelhantes quando *priors* não-informativos foram utilizados e as populações de tamanho grande, de modo geral, apresentaram melhores estimativas. Para as populações pequenas, as análises realizadas considerando os níveis de variabilidade separadamente apresentaram maiores problemas, em virtude do pequeno tamanho das subpopulações formadas. Observou-se, para a metodologia Bayesiana, que o aumento no nível de informação *a priori* influenciou positivamente as estimativas dos componentes de variância, principalmente para as populações pequenas. Portanto, na presença de heterogeneidade de variâncias, as metodologias se comportam de forma semelhante. Entretanto, para populações pequenas a metodologia Bayesiana conduz a melhores estimativas quando informações adicionais estão disponíveis.

Palavras-chave: amostragem de Gibbs, heterocedasticidade, informação *a priori*, melhoramento animal, parâmetros genéticos, simulação

Estimation of variance components using Bayesian and frequentist inferences considering simulated data under heterogeneity of variance

ABSTRACT - A genome of 3.000 centimorgans was simulated for a single quantitative trait governed by 800 *loci* with two alleles per *locus*. According to the genomic structure proposed, 1,500 males and 1,500 females constituted the base population which was used to form two (small and large) initial populations. Two types (only additive genetic variance and both additive genetic and environmental variances) and three variability levels (high, medium and low) of heterogeneity of variances were inserted in the initial populations. Variance components were estimated by Bayesian inference *via* Gibbs Sampling using three different levels of priors (non-informative, slightly informative and informative) and by REML. The best estimates of variance components were obtained with large populations. In small populations, the individual analyses for different variability levels presented problems related to variance components estimation due the small size of subpopulations. Both methods presented similar results for variance components when non-informative priors were used in Bayesian inference. Increasing the level of *a priori* information improved the estimates of variance components by Bayesian inference, especially in small populations.

Key Words: animal breeding, *a priori* information, genetic parameters, Gibbs sampling, heterocedasticity, simulation

Introdução

O principal objetivo do melhoramento animal é o aumento da frequência dos genes favoráveis de determinada

característica dentro da população. Esse processo é realizado por meio da seleção dos melhores indivíduos, que serão utilizados para reprodução. Contudo, a identificação de indivíduos geneticamente superiores tem sido uma

árdua tarefa no melhoramento genético, em que uma das principais dificuldades é quantificar e separar as fontes de variabilidade de uma população.

Com o objetivo de contornar estas dificuldades, inúmeras metodologias estatísticas foram propostas para estimar os componentes de variância e culminaram com a utilização intensiva do Método da Máxima Verossimilhança Restrita (REML). Uma importante pressuposição desta metodologia é que os componentes de variância associados aos efeitos aleatórios sejam constantes entre todas as possíveis estratificações dos dados. No entanto, heterogeneidade de variância genética e residual para rebanhos, produção, região geográfica e tempo, entre outros fatores, tem sido reportada para a maioria das características de interesse econômico.

A heterogeneidade de variância ocorre tanto em características de produção em gado de leite quanto em gado de corte e tem sido amplamente discutida por inúmeros autores, entre eles, Tong et al. (1979), Brotherstone & Hill (1986), Winkelman & Schaeffer (1988), Dong & Mao (1990), Famula (1990), Weigel & Gianola (1992), Kachman & Everett (1993), Meuwissen et al. (1996), Martins (2002), Jensen & Madsen (2002) e Carneiro (2006). A maioria das pesquisas comprova que a heterogeneidade é um problema na própria estrutura dos dados e que pode resultar em classificação errônea dos animais, menor progresso genético e perdas econômicas. Com o objetivo de contornar estas dificuldades, alguns métodos têm sido propostos para considerar a heterogeneidade de variâncias, incluindo transformações, aplicação de fatores de ajustamento e a utilização de modelos multicaracterísticos (Kachman & Everett, 1993).

Alguns autores (Torres et al., 2000; Martins et al., 2000; Cardoso et al., 2001) relataram que as transformações dos dados podem conduzir a pressuposições nem sempre realísticas e que a utilização de fatores multiplicativos, apesar da facilidade computacional, pode produzir avaliações genéticas viesadas.

A abordagem da avaliação genética considerando a heterogeneidade de variâncias, porém usando métodos que a contemplem, permite a estimação dos efeitos fixos de ambiente e a predição dos valores genéticos, ponderando-se adequadamente as observações de acordo com as diferentes variâncias, e pode ser obtida por meio de modelos multicaracterísticos, nos quais a característica que se pretende avaliar é desdobrada em distintos caracteres, um para cada classe de variabilidade, o que permite também considerar a estrutura de covariância.

Contudo, segundo Martins (2002), a desvantagem dessa estratégia está relacionada à necessidade de as classes da heterogeneidade de variância residual serem idênticas às da variância genética aditiva. Adicionalmente, se existirem

muitas classes de variabilidade, a utilização desses modelos pode ser computacionalmente inviável. O pequeno número de animais e as fracas conexões genéticas podem ser os fatores limitantes das análises.

Recentemente, métodos Bayesianos têm sido utilizados como opção para a solução de problemas relacionados à avaliação de mérito genético em populações animais (Gianola & Fernando, 1986). Weigel & Gianola (1992), trabalhando com dados simulados, utilizaram um método Bayesiano para estimação de componentes de variância dentro de rebanho. O método combina informações dentro de rebanhos com estimativas obtidas entre rebanhos. Nesse estudo, foram simulados diferentes tamanhos de rebanhos e diferentes pesos para as informações dentro e entre rebanhos. Os autores notaram que a acurácia da estimativa da variância de touro foi maior quando ambas as informações foram utilizadas: dentro e entre rebanhos. Desse modo, concluíram que a utilização desse método pode permitir a predição de valores genéticos menos influenciados pela heterogeneidade de variâncias.

Nesse contexto, realizou-se este trabalho com os objetivos de verificar o efeito da heterogeneidade de variâncias na estimação dos componentes de variância em populações com dois tamanhos diferentes, comparar o método tradicional de estimação REML com a amostragem de Gibbs que permitiu a inferência Bayesiana e verificar o efeito da introdução de níveis de informação *a priori* nas análises utilizando-se inferência Bayesiana na presença de heterogeneidade de variâncias.

Material e Métodos

Os dados utilizados neste trabalho foram simulados por meio do sistema computacional GENESYS (Euclides, 1996), desenvolvido em linguagem FORTRAN, que permite a simulação de genomas de certa complexidade para formação de populações de indivíduos que podem ser utilizadas em diversos estudos de avaliação genética.

Foi simulado um genoma de 3.000 centimorgans de comprimento, considerando uma única característica quantitativa, governada por 800 locos com dois alelos por loco, na qual a herdabilidade variou de 0,40 a 0,60. Os locos quantitativos foram simulados segundo a distribuição normal e estavam distribuídos em 20 pares de cromossomos de tamanhos aleatórios. Os locos não possuíam desvios de dominância nem interações epistáticas. O genoma não possuía cromossomo sexual e as frequências alélicas iniciais eram iguais em ambos os sexos.

A distribuição matemática usada na simulação das frequências iniciais dos locos quantitativos nas populações

foi a uniforme, com média de 0,50. Utilizou-se um efeito sistemático de rebanho com seis níveis e os efeitos de ambiente não-controláveis foram simulados conforme a distribuição normal. O processo de simulação foi repetido 100 vezes.

Segundo a estrutura genômica proposta, foram simulados 1.500 machos e 1.500 fêmeas para formação da população-base. A partir da população-base foram formadas duas populações iniciais, uma grande e outra pequena. Para formação da população inicial grande, foram escolhidos aleatoriamente 50 machos e 1.500 fêmeas, na razão de um macho para 30 fêmeas, que deram origem a 3.000 descendentes (dois descendentes por fêmea). Essa população foi submetida a uma geração de seleção ao acaso para formação da população inicial grande com 6.000 descendentes. Para formação da população inicial pequena, foram escolhidos aleatoriamente quatro machos e 120 fêmeas, na razão de um macho para 30 fêmeas, que deram origem a 240 descendentes (dois descendentes por fêmea). Essa população foi submetida a uma geração de seleção ao acaso para a formação da população inicial pequena com 480 descendentes.

Com o propósito de avaliar o efeito dos diferentes tipos de heterogeneidade de variâncias, com diferentes tamanhos

de população, e comparar o método freqüentista REML ao Bayesiano via Amostragem de Gibbs (Geman & Geman, 1984), diferentes tipos de estruturas de heterogeneidade foram inseridos nas populações iniciais. Para obtenção destas estruturas de heterogeneidade, foram feitos descartes estratégicos dos valores genéticos e/ou ambientais máximos e mínimos, de acordo com o tipo de heterogeneidade e o nível de variabilidade desejada (Tabela 1).

Em todos os níveis de heterogeneidade estudados, as populações possuíam boa conectividade, obtida pela utilização de touros e vacas em comum entre as populações. Nas populações com heterogeneidade de variância genética aditiva, foram feitos descartes dos valores genéticos reais aditivos máximos e mínimos e, nas populações com heterogeneidade de variância genética aditiva e ambiental, foram feitos descartes dos valores genéticos e ambientais reais máximos e mínimos.

Para ambas as metodologias, foram realizados dois conjuntos de análises. No primeiro conjunto, desconsiderou-se a presença de heterogeneidade de variâncias e admitiu-se que todas as subpopulações pertenciam à mesma população, de modo que as análises foram realizadas com a população total. No segundo conjunto, as análises foram realizadas

Tabela 1 - Código utilizado para as populações com diferentes tipos de heterogeneidade de variâncias, níveis de variabilidade e tamanhos de população

Table 1 - Code used for populations with different types and variability levels of heterogeneity of variance and different size

Estrutura dos dados <i>Data structure</i>	Nível de variabilidade <i>Variability level</i>	Tamanho da população <i>Population size</i>	
		População pequena (P) <i>Small population (P)</i>	População grande (G) <i>Large population (G)</i>
Homogeneidade de variâncias (HO) <i>Homogeneity of variance (HO)</i>	População 1 (1) <i>Population 1 (1)</i>	HO-P-1	HO-G-1
	População 2 (2) <i>Population 2 (2)</i>	HO-P-2	HO-G-2
	População 3 (3) <i>Population 3 (3)</i>	HO-P-3	HO-G-3
	Total (T)	HO-P-T	HO-G-T
Heterogeneidade de variância genética (HG) <i>Heterogeneity of genetic variance (HG)</i>	All (T)		
	Alta (A) <i>High (A)</i>	HG-P-A	HG-G-A
	Média (M) <i>Medium (M)</i>	HG-P-M	HG-G-M
	Baixa (B) <i>Low (B)</i>	HG-P-B	HG-G-B
	Todas (T)	HG-P-T	HG-G-T
Heterogeneidade de variância genética e ambiental (HGA) <i>Heterogeneity of genetic and environmental variances (HGA)</i>	All (T)		
	Alta (A) <i>High (A)</i>	HGA-P-A	HGA-G-A
	Média (M) <i>Medium (M)</i>	HGA-P-M	HGA-G-M
	Baixa (B) <i>Low (B)</i>	HGA-P-B	HGA-G-B
	Todas (T)	HGA-P-T	HGA-G-T
	All (T)		

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

isoladamente para cada subnível de variabilidade. O seguinte modelo linear misto unicaracterístico foi empregado nos dois conjuntos de análises:

$$y_{ij} = \mu + r_i + a_{ij} + e_{ij},$$

em que y_{ij} = observação fenotípica do $j^{\text{ésimo}}$ animal no $i^{\text{ésimo}}$ rebanho; μ = constante inerente a todas as observações; r_i = efeito do $i^{\text{ésimo}}$ rebanho; a_{ij} = efeito aleatório genético aditivo do $j^{\text{ésimo}}$ animal pertencente ao $i^{\text{ésimo}}$ rebanho; e_{ij} = efeito aleatório residual.

Esse modelo pode ser descrito sob a forma matricial a seguir, pressupondo-se que $a \sim N(0, A\sigma_a^2)$ e $e \sim N(0, I\sigma_e^2)$

$$y = X\hat{a} + Za + e$$

em que y = vetor das observações dos indivíduos; X = matriz de incidência de efeitos *fixos*; \hat{a} = vetor de efeitos *fixos* de rebanho; Z = matriz de incidência dos efeitos aleatórios; a = vetor de efeitos aleatórios (valores genéticos dos indivíduos); e e = vetor de erros aleatórios.

As estimativas dos componentes de variâncias pelo método frequentista foram obtidas utilizando-se o programa MTDFREML (*Multiple Trait Derivative-Free Restricted Maximum Likelihood*), descrito por Boldman et al. (1995). Para o método Bayesiano, as estimativas foram obtidas utilizando-se o programa MTGSAM (*Multiple Trait Gibbs Sampling in Animal Models*), descrito por Van Tassel & Van Vleck (1995). Para as análises realizadas pela metodologia Bayesiana, foram utilizados três níveis de informações *a priori*: *priors* não-informativos (PNI), *priors* pouco informativos (PPI) e *priors* informativos (PI). Esses níveis de informações foram obtidos pela variação nos hiperparâmetros da distribuição de qui-quadrado escalonada invertida utilizada para os componentes de variância.

O critério de convergência e as análises das cadeias de Gibbs foram baseados na inspeção visual dos gráficos do valor de cada amostra e no programa GIBANAL (Vankaan, 1998).

Para avaliação dos valores dos componentes de variância, estimados pelas metodologias REML e Bayesiana, calculou-se a porcentagem de erro, por meio da seguinte expressão:

$$PE = \frac{(\hat{s}_i^2 - s_i^2)100}{s_i^2}, \quad i = a, e,$$

em que PE = porcentagem de erro na estimação do componente de variância; \hat{s}_i^2 = componente de variância estimado, s_i^2 = componente de variância real; e $i = a, e$ refere-se às variâncias genéticas aditivas e residuais, respectivamente. Quanto mais próximos de zero os valores de PE mais acurada é a estimativa.

Resultados e Discussão

Os valores reais dos componentes de variâncias genéticas aditivas e ambientais, os valores de herdabilidades e o número de indivíduos simulados (Tabela 2) foram próximos de uma situação real para a característica peso ao sobreano em rebanhos da raça Nelore, como descrito por Balieiro (2001).

Para a estrutura de população com homogeneidade de variâncias utilizada como população controle, esforçou-se para que todos os componentes fossem similares e que as três subpopulações formassem apresentassem herdabilidade próximas entre si.

Tabela 2 - Valores reais dos componentes de variância genética aditiva (\hat{S}_a^2) e ambiental (\hat{S}_e^2) e das herdabilidades (\hat{h}^2) das populações pequenas (P) e grandes (G) com homogeneidade de variâncias (HO), heterogeneidade de variância genética (HG) e heterogeneidade de variâncias genética e ambiental (HGA)

Table 2 - True values of additive genetic (\hat{S}_a^2) and environmental (\hat{S}_e^2) variances and heritability (\hat{h}^2) estimates in small (P) and large (G) populations, with homogeneity of variance (HO), heterogeneity of genetic variance (HG) and heterogeneity of environmental and genetic variances (HGA)

População Population	\hat{S}_a^2	\hat{S}_e^2	\hat{h}^2	Tamanho da população Population size
Homogeneidade de variância Homogeneity of variance				
HO-P-1	179,01	251,02	0,42	100
HO-P-2	182,52	245,29	0,43	100
HO-P-3	188,92	253,71	0,43	100
HO-P-T	189,32	251,46	0,43	300
HO-G-1	303,98	257,65	0,54	1.600
HO-G-2	309,22	236,46	0,57	1.600
HO-G-3	318,21	240,93	0,57	1.600
HO-G-T	310,55	245,04	0,56	4.800
Heterogeneidade de variância genética Heterogeneity of genetic variance				
HG-P-A	216,27	252,42	0,46	100
HG-P-M	159,21	232,31	0,41	100
HG-P-B	125,80	250,00	0,33	100
HG-P-T	171,75	246,00	0,33	300
HG-G-A	303,98	257,65	0,54	1.600
HG-G-M	191,77	237,15	0,45	1.600
HG-G-B	140,46	252,15	0,36	1.600
HG-G-T	212,05	249,01	0,46	4.800
Heterogeneidade de variâncias genética e ambiental Heterogeneity of genetic and environmental variances				
HGA-P-A	259,29	1.302,90	0,17	100
HGA-P-M	135,76	981,54	0,13	100
HGA-P-B	96,93	460,83	0,17	100
HGA-P-T	165,39	921,62	0,15	300
HGA-G-A	339,99	1.414,14	0,19	1.600
HGA-G-M	194,26	868,05	0,18	1.600
HGA-G-B	123,56	476,68	0,21	1.600
HGA-G-T	219,47	919,85	0,19	4.800

Com o intuito de verificar o efeito da heterogeneidade de variâncias nas estimativas obtidas pela metodologia REML, primeiramente desconsiderou-se sua presença, de modo que as análises realizadas com as populações totais referiram-se a cada estrutura de heterogeneidade estudada.

Para as populações pequenas, em todas as estruturas analisadas, os valores de variância genética aditiva foram subestimados quanto ao valor real da população total (Tabela 3). Para a população com homogeneidade de variâncias e tamanho grande, as estimativas foram mais próximas da situação real.

O efeito de se desconsiderar a heterogeneidade de variâncias foi avaliado por meio de estimativas dos componentes de variâncias obtidas a partir da população total, comparadas aos valores reais dos componentes, conforme o nível de variabilidade de interesse (alto, médio ou baixo). A presença da heterogeneidade genética aditiva (Figura 1) afetou a estimação da variância genética aditiva e a heterogeneidade genética aditiva e ambiental afetou ambas as variâncias, genética aditiva e ambiental. Esse fato foi potencializado para as populações pequenas.

As populações com variabilidade média foram menos afetadas pela presença de heterogeneidade (Figura 1), provavelmente porque estas populações possuem valores de variâncias próximos à média das variâncias das três populações. Para as populações grandes, não ocorreram problemas de convergência e as estimativas foram mais acuradas. Entretanto, quando a heterogeneidade não foi considerada, alguns níveis de variabilidade tiveram seus valores subestimados ou superestimados (Figura 1c) e as subpopulações com níveis de variabilidade alta e baixa tiveram suas estimativas de variância genética aditiva subestimadas e superestimadas, respectivamente.

Quando as análises foram realizadas dentro dos níveis (Tabela 3), as populações de tamanho grande, independentemente do tipo de heterogeneidade estudado, apresentaram estimativas mais próximas dos valores reais. Por outro lado, as populações pequenas, em alguns casos, apresentaram problemas na convergência ou as estimativas produzidas foram inconsistentes, em virtude da redução no número de indivíduos quando as análises foram realizadas isoladamente em cada nível de estratificação e quando os componentes foram calculados dentro de cada subclasse de variabilidade. Essa redução no número de indivíduos pode ter causado um problema na própria estrutura dos dados, afetando sua normalidade e dificultando o processo de estimação pela metodologia empregada.

Para as subpopulações HO-G-1, HO-G-2 e HO-G-3 (Tabela 3), as estimativas obtidas (268,86; 88,02 e 91,88) ficaram abaixo dos valores simulados para as subpopulações

Tabela 3 - Desvios das estimativas dos componentes de variâncias genética aditiva (\hat{S}_a^2) e ambiental (\hat{S}_e^2) e valores de herdabilidade (\hat{h}^2) das populações pequenas (P) e grandes (G) com homogeneidade de variâncias (HO), heterogeneidade genética (HG) e heterogeneidade genética e ambiental (HGA)

Table 3 - Differences of additive genetic (\hat{S}_a^2), environmental variance component (\hat{S}_e^2) and heritability (\hat{h}^2) estimates between small (P) and large (G) populations under homogeneity of variance (HO), heterogeneity of genetic variance (HG) and heterogeneity of genetic and environmental variances (HGA)

População Population	\hat{S}_a^2	\hat{S}_e^2	\hat{h}^2	Nº de indivíduos N. of individuals
Homogeneidade de variância Homogeneity of variance				
HO-P-1	FC	FC	FC	100
HO-P-2	109,06	-67,99	0,62	100
HO-P-3	69,58	-17,64	0,52	100
HO-P-T	-125,03	114,89	0,15	300
HO-G-1	-34,81	39,92	0,47	1.600
HO-G-2	-7,33	29,83	0,53	1.600
HO-G-3	53,45	-44,96	0,65	1.600
HO-G-T	60,18	-35,21	0,64	4.800
Heterogeneidade de variância genética Heterogeneity of genetic variance				
HG-P-A	310,66	-197,02	0,90	100
HG-P-M	49,22	5,67	0,47	100
HG-P-B	FC	FC	FC	100
HG-P-T	-127,02	117,26	0,11	300
HG-G-A	-35,12	40,03	0,47	1.600
HG-G-M	-103,75	116,21	0,20	1.600
HG-G-B	-48,58	125,89	0,24	1.600
HG-G-T	-10,26	25,56	0,42	4.800
Heterogeneidade de variâncias genética e ambiental Heterogeneity of genetic and environmental variances				
HGA-P-A	-187,83	88,07	0,05	100
HGA-P-M	FC	FC	FC	100
HGA-P-B	-30,97	13,13	0,12	100
HGA-P-T	-147,45	131,68	0,02	300
HGA-G-A	-83,44	104,5	0,14	1.600
HGA-G-M	-37,92	75,49	0,14	1.600
HGA-G-B	-55,57	54,97	0,11	1.600
HGA-G-T	-58,66	78,41	0,14	4.800

FC = Falhou a convergência.

FC = Convergence failed.

com alta, média e baixa variabilidade (303,98; 191,77 e 140,46, respectivamente). Entretanto, as estimativas de herdabilidade (0,47; 0,53 e 0,65) ficaram próximas das simuladas (0,54; 0,57 e 0,57) (Tabela 2). Para as demais subpopulações oriundas das populações com heterogeneidades de variâncias, as estimativas de herdabilidade foram menos acuradas. Esses resultados são semelhantes aos das análises conduzidas com a população total, na qual, na maioria dos casos estudados, a variância genética aditiva foi subestimada.

As análises realizadas dentro de cada nível de variabilidade não foram eficientes em eliminar os problemas de heterogeneidade sobre a estimação dos componentes de variância quando as subclasses formadas eram pequenas (Tabela 3). Segundo Weigel & Gianola (1992), quando a heterogeneidade existe em rebanhos pequenos, a utilização

de métodos exigentes em normalidade, como o método REML, não produz estimativas confiáveis para componentes de variância dentro de rebanhos. Os resultados também corroboram as observações de Winkelman & Schaeffer (1988) ao demonstrarem que a acurácia das estimativas dos componentes de variâncias é reduzida quando as análises

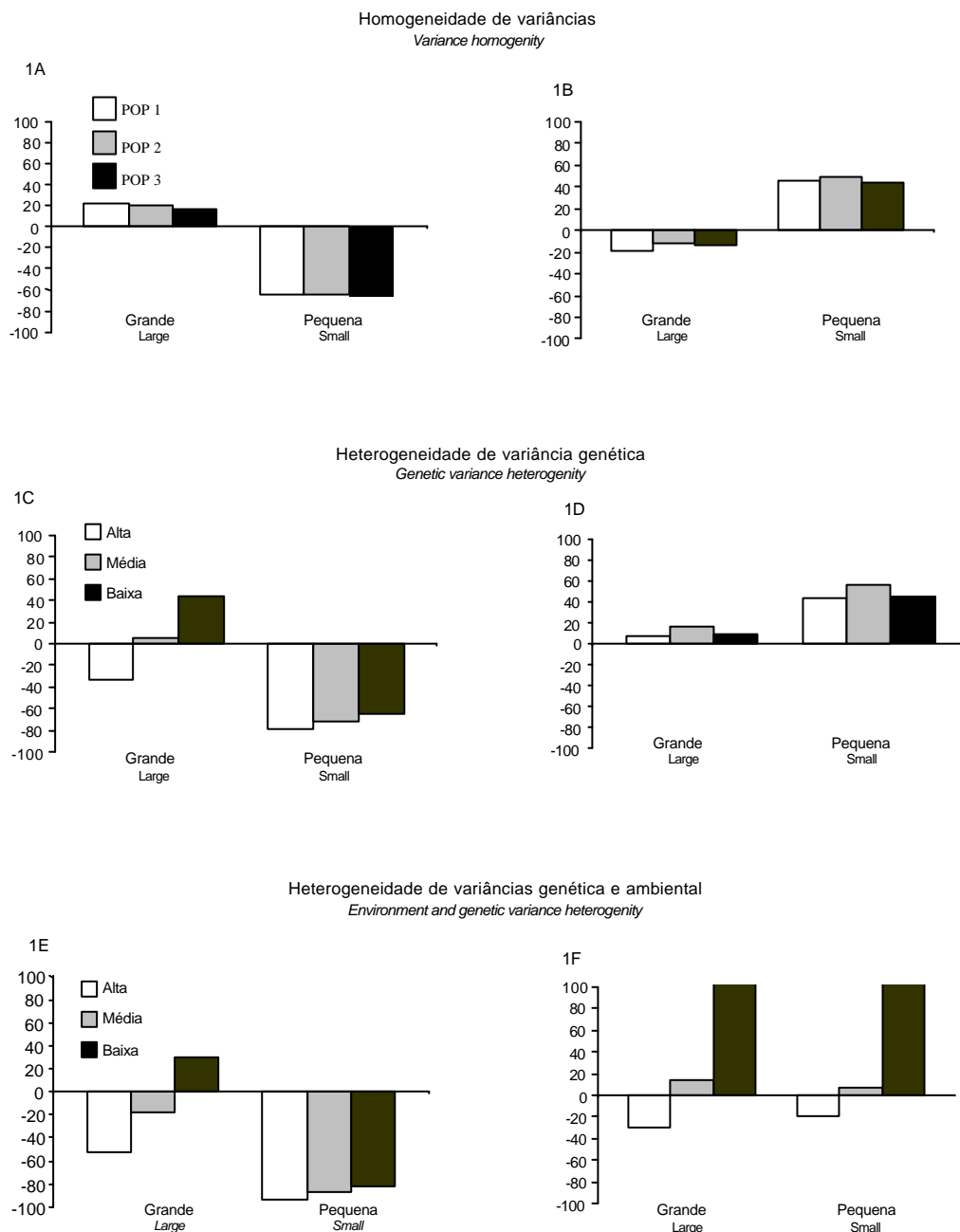


Figura 1 - Porcentagens de erro das estimativas dos componentes de variância genética aditiva (lado esquerdo) e ambiental (lado direito) obtidas pela metodologia REML em populações com 300 indivíduos (PEQUENA) e com 4.800 indivíduos (GRANDE), com homogeneidade de variâncias, heterogeneidade de variância genética e heterogeneidade de variâncias genética aditiva e ambiental, com três níveis de variabilidade.

Figure 1 - Error percentages of additive genetic (left side) and environmental (right side) variance estimates obtained by REML in SMALL (300 individuals) and LARGE (4.800 individuals) populations under homogeneity of variance and three levels of additive genetic variance heterogeneity and environmental and additive variance heterogeneity.

Tabela 4 - Valores da moda das distribuições *a posteriori* dos componentes de variâncias genética aditiva e ambiental das populações pequenas (P) e grandes (G) com homogeneidade de variâncias (HO), heterogeneidade genética (HG) e heterogeneidade genética e ambiental (HGA) estimados sob diferentes níveis de informação *a priori* pela metodologia Bayesiana

Table 4 - Mode values for *a posteriori* distributions of additive genetic and environmental variance in small (P) and large (G) populations under homogeneity of variance (HO), heterogeneity of genetic variance (HG) and heterogeneity of genetic and environmental variances (HGA), estimated using different prior information levels by Bayesian methodology

População <i>Population</i>	$\hat{\sigma}_a^2$			$\hat{\sigma}_e^2$			Número de indivíduos <i>Number of individuals</i>
	PNI	PPI	PI	PNI	PPI	PI	
Homogeneidade de variâncias <i>Homogeneity of variance</i>							
HO-P-1	84,37	84,37	163,57	290,24	290,24	254,38	100
HO-P-2	0,33	165,68	181,71	378,06	275,50	239,11	100
HO-P-3	554,35	119,76	181,99	0,41	266,95	248,11	100
HO-P-T	37,10	87,50	168,34	358,50	340,75	274,67	300
HO-G-1	266,41	277,56	286,21	302,99	296,59	286,80	1600
HO-G-2	318,69	308,65	315,14	275,97	254,70	248,86	1600
HO-G-3	370,71	379,75	334,04	192,23	203,72	223,91	1600
HO-G-T	372,92	373,00	357,15	204,22	212,32	221,88	4800
Heterogeneidade de variâncias genéticas <i>Heterogeneity of genetic variance</i>							
HG-P-A	546,82	196,13	211,99	0,27	245,07	247,49	100
HG-P-M	0,80	78,58	156,01	0,47	303,61	240,31	100
HG-P-B	0,57	49,37	110,03	298,61	245,91	233,15	100
HG-P-T	3,02	85,59	143,63	370,26	327,64	283,74	300
HG-G-A	267,78	279,33	285,81	272,68	289,55	289,49	1600
HG-G-M	95,14	107,22	299,57	329,77	354,35	299,57	1600
HG-G-B	109,68	98,38	120,94	276,20	275,49	258,91	1600
HG-G-T	206,92	215,28	213,83	267,87	266,69	266,14	4800
Heterogeneidade de variâncias genética e ambiental <i>Heterogeneity of genetic and environmental variances</i>							
HGA-P-A	1,30	113,70	190,60	1.313,95	1.161,93	1.251,26	100
HGA-P-M	1,16	62,46	121,64	1.151,70	995,94	1.007,94	100
HGA-P-B	0,77	52,82	85,07	489,10	445,00	451,13	100
HGA-P-T	0,35	66,54	142,21	984,49	993,41	956,32	300
HGA-G-A	282,03	304,65	313,26	1.481,36	1.502,36	1.448,90	1.600
HGA-G-M	183,88	167,01	175,51	940,51	908,87	901,38	1.600
HGA-G-B	76,49	87,01	113,53	510,59	521,25	501,83	1.600
HGA-G-T	169,42	184,33	196,33	978,98	985,59	982,80	4.800

PNI = *Prior* não-informativo; PPI = *Prior* pouco informativo; PI = *Prior* informativo.
PNI = *Non informative priors*; PPI = *Slightly informative priors*; PI = *Informative priors*.

são realizadas dentro dos níveis de variabilidade do rebanho. Segundo esses autores, o pequeno número de animais por rebanho poderia conduzir a grandes erros de amostragem. Além disso, a convergência poderia não ocorrer para alguns rebanhos.

Para melhor visualização (Tabela 4) do efeito de se desconsiderar a heterogeneidade de variâncias, as estimativas das modas das distribuições *a posteriori* dos componentes de variâncias obtidas a partir da população total foram comparadas aos valores reais do nível de interesse por meio das porcentagens de erros.

A introdução de maiores níveis de informação para as populações com homogeneidade de variâncias (Figura 2) resultou em menores porcentagens de erros, tanto para a variância genética aditiva quanto para a variância ambiental.

Entretanto, para as estruturas de população com heterogeneidade genética, o aumento no nível de informação *a priori* somente melhorou a estimação do componente de variância não-heterogêneo (ambiental). Este resultado era esperado, uma vez que, ao desconsiderar a presença de heterogeneidade, apenas uma estimativa é obtida quando, na realidade, existem três situações diferentes: alta, média e baixa variabilidade.

O melhor resultado foi obtido na população grande e com homogeneidade de variâncias (Figuras 2c e 2d). Nesta situação, o aumento no nível de informação *a priori* não resultou em ganhos expressivos de acurácia como para a população pequena (Figuras 2a e 2b), demonstrando que a importância da informação *a priori* aumenta em populações de tamanho pequeno.

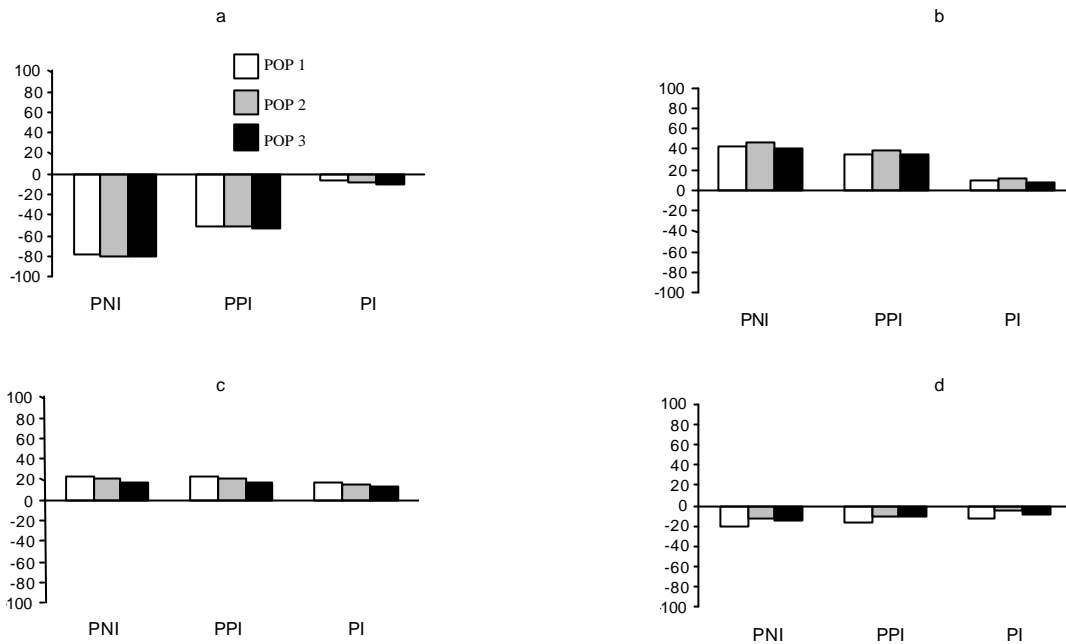


Figura 2 - Porcentagens de erro das estimativas dos componentes de variâncias genética aditiva (2a, 2c) e ambiental (2b, 2d) obtidos pela metodologia Bayesiana sob três níveis de informação *a priori*: não-informativo (PNI); pouco informativo (PPI) e informativo (PI), em populações com 300 indivíduos (9a, 9b) ou com 4.800 indivíduos (2c, 2d), com homogeneidade de variâncias.

Figure 2 - Error percentages for additive genetic (2a, 2c) and environmental (2b, 2d) variance estimates obtained by Bayesian methodology under three *a priori* information levels: Non-informative (PNI), Slightly informative (PPI) and Informative (PI), in populations with 300 individuals (2a, 2b) and 4,800 individuals (2c, 2d), under homogeneity of variance.

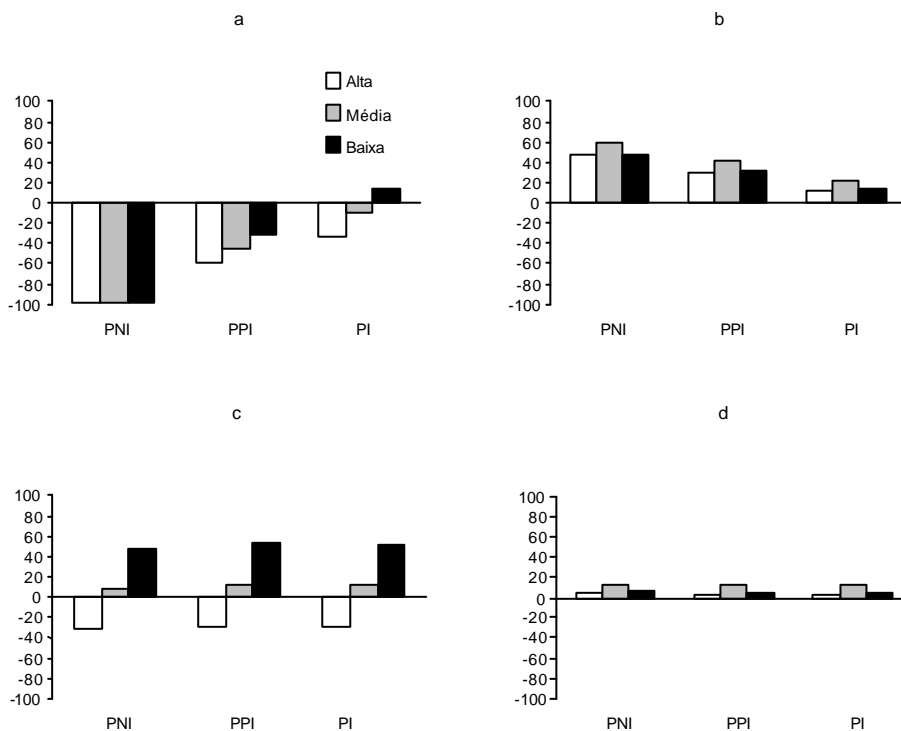


Figura 3 - Porcentagens de erro das estimativas dos componentes de variâncias genética aditiva (3a, 3c) e ambiental (3b, 3d) obtidos pela metodologia Bayesiana sob três níveis de informação *a priori*: não-informativo (PNI); pouco informativo (PPI) e informativo (PI), em populações com 300 indivíduos (3a, 3b) ou com 4.800 indivíduos (3c, 3d), com heterogeneidade de variância genética aditiva com três níveis de variabilidade.

Figure 3 - Error percentages for additive genetic (3a, 3c) and environmental (3b, 3d) variance estimates obtained by Bayesian methodology under three *a priori* information levels: Non-Informative (PNI), Slightly Informative (PPI) and Informative (PI), in populations with 300 individuals (3a, 3b) and 4,800 individuals (3c, 3d), under three variability levels of heterogeneity of genetic variance.

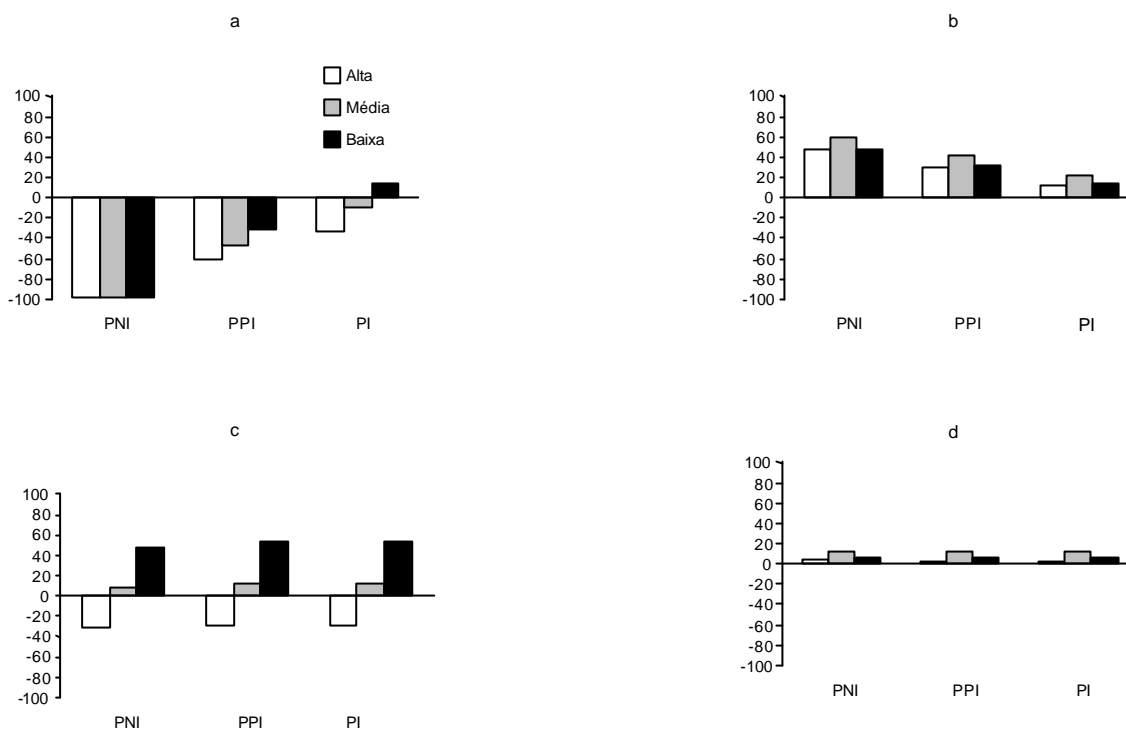


Figura 4 - Porcentagens de erro das estimativas dos componentes de variâncias genética aditiva (14a, 14c) e ambiental (4b, 4d) obtidos pela metodologia Bayesiana sob três níveis de informação *a priori*: não informativo (PNI); pouco informativo (PPI) e informativo (PI), em populações com 300 indivíduos (4a, 4b) ou com 4.800 indivíduos (4c, 4d), com heterogeneidade de variâncias genética aditiva e ambiental, com três níveis de variabilidade.

Figure 4 - Error percentages for additive genetic (4a, 4c) and environmental (4b, 4d) variance estimates obtained by Bayesian methodology, under three *a priori* levels of information: Non-Informative (PNI), Slightly Informative (PPI) and Informative (PI), in populations with 300 individuals (4a, 4b) and 4,800 individuals (4c, 4d), under three variability levels of heterogeneity of genetic and environmental variances.

Na situação em que a presença da heterogeneidade genética aditiva foi desconsiderada (Figura 3), até mesmo a população de tamanho grande (Figura 3C) apresentou problemas na estimação da variância genética aditiva, pois o nível de informação *a priori* não foi capaz de melhorar a estimação. Entretanto, a variância ambiental não foi influenciada pela heterogeneidade de variância genética aditiva.

As subpopulações com variabilidade média não apresentaram problemas pelo fato de seus valores de variância ambiental estarem próximos do valor médio dos três níveis. Na estrutura de população apresentando heterogeneidade nos dois componentes de variância (Figura 4), tanto a variância genética aditiva (Figuras 4A e 4B) quanto a variância ambiental (Figura 4C e 4D) tiveram seus valores superestimados para o nível de alta variabilidade e subestimados para o nível de baixa variabilidade. Nesta situação, a introdução de níveis de informação *a priori* não produziu efeito satisfatório, com exceção da população pequena (Figura 4A), na qual a convergência não foi obtida quando se utilizou *prior* não-informativo.

Para as populações pequenas em todos os tipos de heterogeneidade estudados e em alguns níveis de variabili-

dade, quando *priors* não-informativos foram utilizados, à semelhança do método REML, ocorreram problemas na convergência ou as estimativas de componentes de variâncias obtidas não foram consistentes (Tabela 4). Entretanto, quando utilizados maiores níveis de informação, as estimativas se aproximaram dos valores reais. Esses resultados estão de acordo com os obtidos por Weigel & Gianola (1992), que concluíram que a utilização da informação *a priori* pode contribuir para a estimação dos componentes de variância. De modo geral, as estimativas de componentes de variâncias mais acuradas foram obtidas nos níveis mais informativos (PI e I), semelhantemente aos resultados descritos por diversos autores (Wang et al. 1994; Blasco et al. 1998). Segundo Weigel & Gianola (1992) à medida que o tamanho da amostra aumenta, a distribuição *a posteriori* se assemelha à distribuição representada pela função de verossimilhança, demonstrando que o conhecimento *a priori* tende a perder a importância relativa com o aumento do volume dos dados das populações.

Como previsto, as estimativas produzidas pelo método REML e pelo método Bayesiano com *priors* não-informa-

tivos foram semelhantes. A vantagem do método Bayesiano, neste caso, foi a utilização de informações *a priori* nas análises realizadas dentro dos níveis, principalmente em subpopulações de tamanho pequeno, indicando que a informação *a priori* é relativamente mais importante quando os dados são escassos (Gianola & Fernando, 1986).

Segundo Weigel & Gianola (1992), o conhecimento deste fato pode ser utilizado para estimação de componentes de variância combinando informações obtidas entre e dentro de rebanhos, que resultam em estimativas menos influenciadas pela heterogeneidade de variâncias.

Conclusões

A desconsideração dos diferentes tipos de heterogeneidade de variâncias reduz a acurácia nas estimativas dos componentes de variâncias. O aumento do nível de conhecimento da informação *a priori* conduz a melhores estimativas de componentes de variância no método Bayesiano, principalmente para populações pequenas.

Na ausência de *priors* informativos, os métodos REML e Bayesiano apresentam resultados semelhantes. Na presença de heterogeneidade de variâncias, as estimativas das variâncias são subestimadas e superestimadas nos níveis de alta e baixa variabilidade, respectivamente, e níveis de variabilidade média são menos prejudicados. A presença de heterogeneidade de variâncias em apenas um dos componentes, genético aditivo ou ambiental, não afeta a estimação do componente homogêneo. A heterogeneidade nos dois componentes conduz a maiores prejuízos no processo de estimação.

O método Bayesiano é indicado para populações de tamanho pequeno quando há disponibilidade de *priors* informativos.

Literatura Citada

- BALIEIRO, J.C.C. **Heterogeneidade de variância na avaliação genética de bovinos da raça Nelore**. Viçosa: Universidade Federal de Viçosa, 2001. 88p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 2001.
- BLASCO, A.; SORENSEN, D.; BIDANEL J.P. Bayesian inference of genetic parameters and selection response for litter size components in pigs. **Genetics**, v.149, p.301-306, 1998.
- BOLDMAN, K.G.; KRIESE, L.A.; van VLECK, L.D. et al. **A manual for use of MTDFREML: a set of programs to obtain estimates of variances and covariances (DRAFT)**. Lincoln: Department of Agriculture/Agriculture Research Service, 1995. 120p.
- BROTHERSTONE, S.; HILL, W.G. Heterogeneity of variance amongst herds for milk production. **Animal Production**, v.42, n.3, p.297-303, 1986.
- CARDOSO, F.F.; CARDELLINO, R.A.; CAMPOS, L.T. Fatores ambientais sobre escores de avaliação visual à desmama em bezerras Angus criados no Rio Grande do Sul. **Revista Brasileira de Zootecnia**, v.30, n.2, p.318-325, 2001.
- CARNEIRO, A.P.S.; TORRES, R.A.; LOPES, P.S. Erros na classificação de touros, vacas e touros jovens geneticamente superiores avaliados na presença de heterogeneidade. **Revista Brasileira de Zootecnia**, v.35, n.5, p.1907-1914, 2006.
- DONG, M.C.; MAO, I.L. Heterogeneity of (co)variance and heritability in different levels of intraherd milk production variance and of herd average. **Journal of Dairy Science**, v.73, p.843-85, 1990.
- EUCLYDES, R.F. **Uso do sistema para simulação Genesys na avaliação de métodos de seleção clássicos e associados a marcadores moleculares**. Viçosa, MG: Universidade Federal de Viçosa, 1996. 150p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 1996.
- FAMULA, T.R. Detection of heterogeneous variance in herd production groups. **Journal of Dairy Science**, v.72, n.3, p.715-721, 1990.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.6, p.721-741, 1984.
- GIANOLA, D.; FERNANDO, L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217-244, 1986.
- JENSEN, J.; MADSEN, P., **Bayesian estimation of within and across country genetic parameters for MACE Tjele Kenmark**: Danish Institute of Agricultural Sciences, 2002 (CD-ROM).
- KACHAMN, S.D.; EVERETT, R.W. A multiplicative mixed model when the variances are heterogeneous. **Journal of Dairy Science**, v.76 n.3, p.859-867, 1993.
- MARTINS, E.N.; SILVA, M.A.; ALENCAR, M.M. et al. Uso de fatores multiplicativos para ajustamento de dados. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 37., 2000, Viçosa, MG. **Anais...** Viçosa, MG: Sociedade Brasileira de Zootecnia, 2000. (CD-ROM).
- MARTINS, E.N. Avaliação genética e heterogeneidade de variância. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 39., 2002, Recife. **Anais...** Recife: Sociedade Brasileira de Zootecnia (2002). (CD-ROM).
- MEUWISSEN, T.H.E.; JONG, G.; ENGEL, B. Joint estimation of breeding values and heterogeneous variances of large data sets. **Journal of Dairy Science**, v.79, n.2, p.310-316, 1996.
- TONG, A.K.W.; KENNEDY, B.W.; MOXLEY, J. E. Potential errors in sire evaluation from regional genetic differences. **Journal of Dairy Science**, v.63, n.4, p.627-633, 1979.
- TORRES, R.A.; BERGMANN, J.A.G.; COSTA, C.N. et al. Heterogeneidade de variância e avaliação genética de bovinos da raça Holandesa no Brasil. **Revista Brasileira de Zootecnia**, v.29, n.4, p.1050-1059, 2000.
- Van TASSEL, C.P.; Van VLECK, L.D. **A manual for use of MTGSAM. A set of Fortran programs to apply Gibbs Sampling to animal models for variance components estimation**. Lincoln: United States Department of Agriculture, Agriculture Research Service, 1995. 85p.
- Van KAAM, J.B.C.H.M. "GIBANAL" - Analyzing program for Markov Chain Monte Carlo Sequences (Version 2.10). Wageningen: Department of Animal Sciences, Wageningen Agricultural University, 1998. 4p. (Manual).
- WANG, C.S.; GIANOLA, D.; SORENSEN, D.A. et al. Response to selection for litter size in danish landrace pigs: a bayesian analysis. **Theory Applied Genetics**, v.88, p.220-230, 1994.
- 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.
- WINKELMAN, A.; SCHAEFFER, S.R. Effect of heterogeneity of variance on dairy sire evaluation. **Journal of Dairy Science**, v.71, p.3033-3039, 1988.

Recebido: 12/6/2006
Aprovado: 26/4/2007