

JOSÉ MARQUES CARNEIRO JÚNIOR

**AVALIAÇÃO DA HETEROGENEIDADE DE VARIÂNCIAS UTILIZANDO
DADOS SIMULADOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de “Doctor Scientiae”.

VIÇOSA
MINAS GERAIS - BRASIL
2005

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

C289a
2005

Carneiro Júnior, José Marques, 1974-
Avaliação da heterogeneidade de variâncias utilizando
dados simulados / José Marques Carneiro Júnior.- Viçosa
: UFV, 2005.
xi, 88f. : il. ; 29cm.

Orientador: Ricardo Frederico Euclides
Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 82-88

1. Melhoramento animal - Métodos de simulação.
2. Análise de variância. 3. Marcadores genéticos. 4. Teoria
bayesiana de decisão estatística. 5. Genética. I. Universi-
dade Federal de Viçosa. II. Título.

CDD 22.ed. 636.0824

JOSÉ MARQUES CARNEIRO JÚNIOR

**AVALIAÇÃO DA HETEROGENEIDADE DE VARIÂNCIAS UTILIZANDO
DADOS SIMULADOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de “Doctor Scientiae”.

APROVADA: 14 de fevereiro de 2005.

Prof. Paulo Sávio Lopes
(Conselheiro)

Prof. Robledo de Almeida Torres
(Conselheiro)

Prof^a. Carmen Silva Pereira

Prof. Ricardo da Fonseca

Prof. Ricardo Frederico Euclides
(Orientador)

Aos meus pais José Marques Carneiro e Maria
Valdise Martins Carneiro.

Aos meus irmãos Sólon, Andréia, Adriana e
Alessandra.

A minha esposa Giselle e a minha filha Gabriela.
Dedico este trabalho.

AGRADECIMENTOS

A Deus.

À Universidade Federal de Viçosa, em especial ao Departamento de Zootecnia e ao Programa de Pós-Graduação em Genética e Melhoramento.

À Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudo.

À professora Carmen Silva Pereira, pelo exemplo de dedicação à pesquisa científica, pelos ensinamentos, pelas sugestões e críticas que muito contribuíram para a finalização deste trabalho.

Ao professor Paulo Sávio Lopes, pela amizade, ensinamentos, críticas e apoio que muito contribuíram para a finalização deste trabalho.

Ao professor Ricardo Frederico Euclides, pela disponibilização do programa de simulação Genesys, pela orientação e amizade que foram fundamentais para a elaboração deste trabalho.

Ao professor Robledo de Almeida Torres, pela amizade, pelos ensinamentos, pelas sugestões sempre coerentes, pelos incentivos que muito contribuíram para a execução deste trabalho.

Aos demais professores do Programa de Pós-Graduação em Genética e Melhoramento, pelos ensinamentos e atenção.

Ao professor Henrique Nunes de Oliveira, pela receptividade e concessão de alguns programas que contribuíram para a conclusão deste trabalho.

Aos meus pais, por terem me dado o maior de todos os presentes – o estudo. Por terem assumido o compromisso de educar com

responsabilidade, sinceridade e amor, deixando aos filhos uma grande herança, a educação.

As minhas irmãs Andréa, Adriana e Alessandra que oportunamente me incentivaram na realização dos meus objetivos.

Ao meu irmão e primo Sólon pela alegria, pelos momentos de descontração e pelo incentivo a minha formação.

A minha esposa Giselle pelo incentivo, pelas dicas de grande utilidade, pela dedicação e carinho, que fizeram deste trabalho uma realização agradável.

A minha filha Gabriela por fazer parte da minha vida.

A minha Sogra Tânia Ludmila e ao meu Sogro Fernando, pelo incentivo, amizade e apoio.

Aos colegas e amigos de curso, Adriana, Audrin, Amauri, Claudinho, Elizângela, Fausto, Felipe, Fernanda, Gustavo, Jaime, Leandro, Lindenberg, Paulo Nobre, Paulo Carneiro, Rachel, Fábio, Jane, Ricardinho, Rodolpho, Urbano e Willian, pela amizade, companheirismo e pelo convívio agradável.

Aos amigos Barnabé, Juninho, Pedro Malazarte, Lamina Oterp, Areia da Silva e Walter pelos momentos felizes e descontraídos.

As amigas Daniela, Laiza, Elizângela, Lili e Adriana, pela amizade agradável e sincera.

Aos professores e funcionários do Departamento de Zootecnia e a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

BIOGRAFIA

JOSÉ MARQUES CARNEIRO JÚNIOR, filho de José Marques Carneiro e Maria Valdisé Martins Carneiro, nasceu em Porto Velho, Estado de Rondônia, em 27 de março de 1974.

Em março de 1994, iniciou na Universidade Federal de Viçosa o curso de graduação em Zootecnia, concluído em março de 1999.

Em abril de 1999, ingressou no curso de Mestrado em Genética e Melhoramento na UFV, realizando estudos na área de Melhoramento Animal.

No dia 22 de março de 2001, submeteu-se aos exames de defesa de tese.

Em abril de 2001, iniciou o curso de Doutorado em Genética e Melhoramento na UFV, submetendo-se ao exame final de defesa de tese em 14 de fevereiro de 2005.

ÍNDICE

RESUMO	viii
ABSTRACT	x
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1. Heterogeneidade de Variâncias.....	3
2.2. A Estatística Bayesiana	11
2.2.1. Teorema de Bayes.....	11
2.2.2. Os Métodos Bayesianos	13
2.2.3. O Amostrador de Gibbs	14
2.2.4. Distribuições <i>a priori</i>	16
2.2.5. Distribuições Robustas	17
2.2.5.1. Distribuição Normal Contaminada	17
3. MATERIAL E MÉTODOS	21
3.1. Simulação do genoma	21
3.2. Simulação das Populações.....	22
3.3. Análise dos dados	24
3.3.1. Pressuposições e Distribuições <i>a priori</i>	26
3.3.2. Critério de Convergência	28
3.4. Critérios de avaliação e variáveis avaliadas	28
3.4.1. Acurácia das estimativas das variâncias	28
3.4.2. Correlações entre os valores genéticos.....	29
3.4.3. Quadrado Médio do Erro (QME).....	29
3.4.4. Cálculo dos Componentes de Variâncias Reais	30
4. RESULTADOS E DISCUSSÃO	31
4.1. Valores Reais dos Componentes de Variância.....	31
4.2. Estimação dos componentes de variância.....	34
4.2.1. Análises pela Metodologia REML	34
4.2.2. Análises pela Metodologia BAYESIANA.....	43

4.3. Predição dos valores genéticos	58
4.3.1. Correlação de Ordem – SPEARMAN	58
4.3.2. Quadrado Médio do Erro	64
4.3.3. Porcentagens de Indivíduos Seleccionados em Comum.....	71
4.3.4. Porcentagens de Indivíduos Seleccionados por Ambiente.....	77
5. CONCLUSÕES	81
REFERÊNCIAS BIBLIOGRÁFICAS.....	82

RESUMO

CARNEIRO JÚNIOR, José Marques, D. S., Universidade Federal de Viçosa, fevereiro de 2005. **Avaliação da heterogeneidade de variâncias utilizando dados simulados.** Orientador: Ricardo Frederico Euclides. Conselheiros: Paulo Sávio Lopes e Robledo de Almeida Torres.

Estudos de simulação foram conduzidos com o objetivo de realizar uma análise comparativa, via simulação de dados, entre a metodologia clássica de estimação dos componentes de variância e predição dos valores genéticos REML – BLUP e a metodologia Bayesiana que permite a inclusão de informação *a priori* e a utilização de distribuições robustas, como a normal contaminada, na avaliação genética dos animais. Foi simulado um genoma de 3000 centimorgans de comprimento, considerando uma única característica quantitativa, governada por 800 locos com dois alelos por loco, na qual a herdabilidade variou conforme a estrutura desejada de heterogeneidade de variâncias. Segundo a estrutura genômica proposta, foram simulados 1500 machos e 1500 fêmeas que formaram a população base. A partir da população-base foram formadas duas populações iniciais, uma grande e outra pequena. Com o propósito de avaliar o efeito dos diferentes tipos de heterogeneidade de variâncias, em populações com dois tamanhos, bem como comparar o método REML – BLUP com o método Bayesiano, foram inseridos diferentes tipos de estruturas de heterogeneidade nas populações iniciais. Para obtenção destas estruturas de heterogeneidade foram feitos descartes estratégicos dos valores genéticos, ambientais, ou de ambos, de acordo com o tipo de heterogeneidade e o nível de variabilidade desejada: alta, média ou baixa. Para a metodologia Bayesiana foram utilizados três níveis de informação *a priori*: não informativo, pouco informativo e informativo. Para a estrutura com

heterogeneidade ambiental foi empregado também o método Bayesiano, considerando distribuição normal contaminada para os resíduos. De forma geral foi verificado que a presença da heterogeneidade causa problemas para seleção dos melhores indivíduos, principalmente se a heterogeneidade estiver presente no componente ambiental. Os métodos comparados apresentaram resultados semelhantes quando *priors* não informativos foram utilizados, sendo que as populações de tamanho grande, de modo geral, apresentaram melhores estimativas. Para as populações pequenas as análises realizadas dentro dos subníveis apresentaram maiores problemas, devido ao pequeno tamanho das subpopulações formadas. Foi observado, 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. A utilização da distribuição normal contaminada para os resíduos, não foi eficiente em eliminar os problemas causados pela presença da heterogeneidade de variâncias, sendo que para predição dos valores genéticos os resultados foram similares. Apesar do aumento de informação ter conduzido a estimativas mais acuradas de componentes de variância, a correlação de Spearman entre os valores genéticos reais e preditos não foi alterada quando níveis mais informativos foram utilizados. Contudo, foi verificado pelo Quadrado Médio do Erro que a predição dos valores genéticos foi sensivelmente mais acurada, quando o maior nível de informação foi utilizado. Conclui-se, portanto, que melhores predições dos valores genéticos, para populações pequenas, podem ser obtidas pela metodologia Bayesiana quando informações adicionais estão disponíveis.

ABSTRACT

CARNEIRO JUNIOR, José Marques, D. S., Universidade Federal de Viçosa, february 2005. **Evaluation of variance heterogeneity using simulate data**. Adviser: Ricardo Frederico Euclides. Committee members: Paulo Sávio Lopes and Robledo de Almeida Torres.

Studies on simulation were carried out aiming to achieve a comparative analysis, through data simulation, between the classic methodology REML - BLUP of the variance components estimation and genetic values prediction and the Bayesian methodology, that allows the inclusion of *a priori* information and the use of robust distributions, as the contaminated normal distribution, in the animal genetic evaluation. A genome of 3,000 length centimorgans was simulated, considering a single quantitative trait, governed by 800 loci with two alleles by locus, in which heritability varied accordingly with the heterogeneity variance structures desired. According to the genomic structure proposed, there were simulated 1,500 males and 1,500 females that formed the base population. Starting from the base population, two initial populations were formed: a large and a small one. With the purpose of evaluating different type of heterogeneity variance effects, in populations with two sizes, as well as to compare the method REML - BLUP with the Bayesian method, different types of heterogeneity structures were inserted in the initial populations. For obtaining these heterogeneity structures there were made strategic discards of genetic values, environmental, or both, in agreement with the heterogeneity type and the level of desired variability: high, medium or small. For Bayesian methodology, three *a priori* information levels were used: no informative, slightly informative and informative. For structure with environmental heterogeneity, it was also used the Bayesian method considering contaminated normal distribution for the residuals. In a general way, it was verified that the presence of the heterogeneity causes problems for the best individuals' selection, mainly if the heterogeneity occurs in the environmental component. The compared

methods presented similar results when no informative *priors* were used, and large size populations presented, in general, better estimates. For small populations, the analyses accomplished inside of the subclass presented larger problems, due to small size of the formed subclass. It was observed, for the Bayesian methodology, that the increase the *a priori* information level influenced the estimates of variance components positively, mainly for the small populations. Using contaminated normal distribution for the residues, was not efficient in eliminating the problems caused by variances heterogeneity, and for genetic values prediction the results were similar. In spite of the increase of information to have led to accurate estimates of the variance components, the Spearman Correlation among the true genetic values and predicted was not altered when more informative levels were used. However, it was verified by the Mean Square Error that prediction genetic values was sensibly more accurate, when more information level was used. It is ended, therefore, that better predictions of the genetic values, for small populations, they can be obtained by the Bayesian methodology when additional information are available.

1. INTRODUÇÃO

No melhoramento animal o objetivo principal é o aumento da frequência dos genes favoráveis. Este processo é realizado por meio da escolha dos melhores indivíduos, que serão utilizados para reprodução.

Se toda variação observável nos animais fosse herdável, isto é, fosse de origem genética, a seleção dos melhores indivíduos seria um processo fácil e não haveria necessidade de métodos estatísticos tão sofisticados serem empregados. Entretanto, o melhoramento animal caracteriza-se pela seleção de características de baixa herdabilidade, dados de baixa qualidade e uma grande influência não genética nas comparações entre os indivíduos.

Com o objetivo de contornar estes problemas, inúmeras metodologias estatísticas foram propostas, culminando com a utilização intensiva do método BLUP – Melhor Predição Linear não Viesada, associado ao método REML – Máxima Verossimilhança Restrita, para predição dos valores genéticos e estimação dos componentes de variância, respectivamente.

Uma importante suposição na maioria das aplicações de metodologias de avaliação genética, em particular o BLUP, é que os componentes de variância associados aos efeitos aleatórios sejam constantes entre todas as possíveis estratificações dos dados. Entretanto, heterogeneidade de variância genética e residual para rebanhos, produção, região geográfica e tempo, dentre outros fatores, existem na maioria das características de interesse econômico. 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).

Avaliações genéticas utilizando o BLUP podem propriamente considerar a heterogeneidade de variâncias. Isto é possível com a utilização de modelos multicaracterísticos, onde cada característica que se pretende avaliar é considerada como diferente característica em cada classe de heterogeneidade de variância. Entretanto, quando a heterogeneidade existe em subclasses com poucos dados, métodos que necessitam de justificação assintótica, tal qual o método REML, podem produzir estimativas de componentes de variância não confiáveis dentro dessas subclasses.

Recentemente, métodos bayesianos vêm sendo utilizados como uma opção para a solução de problemas relacionados à avaliação de mérito genético em populações animais. A aplicação de métodos *Markov Chain Monte Carlo* (MCMC), dentre os quais se destaca a Amostragem de Gibbs, pode ser utilizada de forma a propiciar uma inferência Bayesiana. *Gibbs Sampling* (GS) é uma técnica de integração numérica por simulação, muito usual em situações nas quais a integração analítica completa é impossível. O GS é aplicável à estimação de componentes de variância e permite, por suas propriedades, a inferência bayesiana.

Dentro do contexto apresentado, os objetivos desse trabalho foram:

- Verificar o efeito de diferentes tipos de heterogeneidade de variâncias e do tamanho de populações na estimação dos componentes de variância e na predição dos valores genéticos;
- Verificar a utilização da distribuição normal contaminada para os resíduos, em dados com heterogeneidade ambiental;
- Verificar a influência do nível de informação *a priori* ao utilizar metodologia Bayesiana na análise genética; e
- Comparar as metodologias clássicas REML – BLUP com a Bayesiana na análise genética em dados com heterogeneidade de variâncias.

2. REVISÃO DE LITERATURA

2.1. HETEROGENEIDADE DE VARIÂNCIAS

A identificação e o acasalamento de animais geneticamente superiores é essencial para o progresso genético de uma população. Entretanto, para a realização da seleção, necessário se faz o conhecimento dos componentes de variância da população com o qual se está trabalhando. Estimativas dos componentes de variância e coeficientes de herdabilidade têm se mostrado heterogêneas de acordo com diferentes níveis de produção, classes de desvio-padrão genético e/ou ambiental e tamanho do conjunto de dados dos rebanhos.

Quando animais são comparados entre diferentes ambientes, o ajuste para diferenças entre ambientes é um processo importante. Não apenas alguns ambientes são mais favoráveis do que outros, mas a variabilidade também pode diferir. 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, dentre eles TONG et al. (1979); BROTHERSTONE e HILL (1986); FAMULA (1990); WINKELMAN e SCHAEFFER (1988); DONG e MAO (1990); WEIGEL e GIANOLA (1992); KACHMAN e EVERETT (1993); MEUWISSEN et al. (1996); MARTINS (2002); JENSEN e MADSEN (2002) e CARNEIRO (2003). 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. Inúmeros métodos têm sido propostos para considerar a heterogeneidade de variâncias, incluindo transformações, aplicação de fatores de ajustamento e a utilização de modelos multicaracterísticos (KACHMAN e EVERETT, 1993).

Segundo MARTINS (2002), duas linhas básicas de abordagem da avaliação genética, pressupondo a heterogeneidade de variâncias, têm sido seguidas. Uma propõe o uso de métodos que permitam eliminar a heterogeneidade de variância e a outra propõe o uso de métodos que a contemplem. Os métodos que buscam eliminar a heterogeneidade de variâncias consistem na transformação de dados ou na aplicação de fatores de ajustamento de forma que os dados transformados ou ajustados apresentem homogeneidade de variâncias. Diversos autores, entre eles CARDOSO et al. (2001); TORRES et al. (2000) e MARTINS et al. (2000) relatam que as transformações de dados podem conduzir a pressuposições nem sempre realísticas e que a utilização de fatores multiplicativos, apesar da facilidade computacional, podem produzir avaliações genéticas viesadas.

A abordagem da avaliação genética, pressupondo heterogeneidade de variâncias, porém usando métodos que a contemplem, permite a estimação dos efeitos fixos de ambiente e predição dos valores genéticos, ponderando adequadamente as observações de acordo com as diferentes variâncias em cada classe de heterogeneidade e, ainda, considerando a estrutura de covariância.

TONG et al. (1979), trabalhando com dados simulados, avaliaram o efeito de se ignorar diferenças genéticas entre regiões nas avaliações de touros. Os autores concluíram que se não existe troca genética entre rebanhos e se touros utilizados em uma região não são relacionados com touros utilizados em outras regiões, as avaliações de reprodutores devem ser restritas dentro de sua região, e comparações entre regiões não podem ser realizadas.

WINKELMAN e SCHAEFFER (1988) avaliaram o efeito da heterogeneidade de variâncias dentro de rebanhos de gado leiteiro. Neste estudo foi utilizado um modelo multicaracterístico com o objetivo de considerar a presença de heterogeneidade de variâncias. Contudo, não foram encontradas diferenças na acurácia das avaliações de touros entre o modelo convencional e o multicaracterístico. Os autores concluíram que o problema da heterogeneidade de variâncias é maior nas avaliações de vacas dentro de rebanhos.

FAMULA (1990), com o objetivo de detectar a heterogeneidade de variâncias, avaliou o efeito de se estratificar rebanhos leiteiros em níveis de

variabilidade. O autor concluiu que o procedimento de estratificação dos dados pela média da produção leiteira em baixa, média e alta seguido do cálculo das variâncias genética e residual para cada segmento independentemente, pode causar uma mudança na natureza do problema. Neste caso, passa-se de uma situação de variâncias heterogêneas para estimação em dados selecionados.

BOLDMAN e FREEMAN (1990), trabalhando com diferentes níveis de produção: baixa, média e alta, e diferentes modelos com e sem transformação dos dados, concluíram que as avaliações de touros são pouco afetadas pela heterogeneidade de variâncias. Entretanto, a avaliação de vacas é mais afetada pela violação de se assumir estrutura de homogeneidade de variâncias no modelo. Os autores destacam também que a transformação logarítmica não é recomendada porque vacas criadas em rebanhos de baixa produção tendem a ter seus valores genéticos superestimados.

DONG e MAO (1990), utilizando dados reais, avaliaram diferentes critérios de estratificação dos dados, para identificação de variâncias heterogêneas entre rebanhos: critério de variância dentro de rebanho-ano; critério de variância dentro de rebanho; critério de média de produção de rebanho-ano; e critério de média de produção de rebanho. Os autores concluíram que a estratificação dos registros por meio do critério de variância dentro de rebanho-ano foi o mais efetivo para identificar a heterogeneidade de variâncias.

Segundo WIGGANS e VanRADEN (1991), vários tipos de correção para heterogeneidade de variâncias tem sido investigadas e implementadas. Entretanto, a maioria dessas metodologias desenvolvidas, principalmente para estimação de variâncias individuais para rebanho-ano, requer mais informações do que usualmente estão disponíveis. Outra desvantagem mencionada pelos autores é que estes métodos assumem que a herdabilidade seja constante entre todos os níveis de variâncias fenotípicas.

WEIGEL e GIANOLA (1992), utilizando dados simulados verificaram a utilização de um método Bayesiano para estimação de componentes de variância dentro de rebanho. O método utilizado 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 de rebanhos e estimativas obtidas entre rebanhos. Os

autores encontraram que a acurácia da estimativa da variância de touro foi maior quando ambas as informações foram utilizadas: dentro e entre rebanhos. Entretanto, o aumento verificado na acurácia da variância residual com incorporação da informação *a priori* foi mínimo. Os autores concluíram que a utilização desse método pode permitir a predição de valores genéticos menos influenciados pela heterogeneidade de variâncias.

KACHMAN e EVERETT (1993) propõem a utilização de um modelo misto multiplicativo, utilizando a inferência Bayesiana, que ignora a relação linear entre variâncias e médias. Segundo os autores, o método proposto funciona bem quando a herdabilidade e a repetibilidade se mantêm constante ao longo das classes de heterogeneidades de variâncias. Pressuposição não realista para a maioria dos conjuntos de dados verificados no melhoramento genético animal.

LEE e PARK (1994) encontraram heterogeneidade de variâncias, em dados de produção de leite em rebanhos da Coréia. Os autores destacam que a maior parte da heterogeneidade é devido a diferenças ambientais entre os rebanhos.

TORRES et al. (2000) trabalhou com heterogeneidade de variâncias na avaliação genética de vacas e touros da raça Holandesa criados no Brasil. A produção total de leite ajustada para idade adulta foi usada para dividir os rebanhos em três classes de desvio-padrão fenotípico: alto, médio e baixo. O autor encontrou que as médias de produção de leite e os componentes de variâncias genética, residual e fenotípica aumentaram com o aumento do desvio-padrão médio da classe, sendo que as herdabilidades não apresentaram o mesmo comportamento. As transformações utilizadas pelo autor para corrigir o problema de heterogeneidade não foram eficientes. Nas análises onde cada ambiente foi considerado como sendo uma característica distinta, as correlações entre os valores genéticos, para as classes de baixo, médio e alto desvio-padrão e os obtidos na análise geral, foram próximos da unidade, indicando que os reprodutores seriam classificados similarmente nas três classes de desvio-padrão. O autor conclui que na avaliação genética de reprodutores, torna-se importante considerar a variabilidade dentro de níveis de produção, pois, sob seleção, as classes mais variáveis contribuiriam com a

maior parte dos animais, e avaliação genética do animal poderia ser mais em função do seu ambiente que do seu potencial genético.

CAMPÊLO (2001), trabalhando com dados reais, avaliou a importância da heterogeneidade de variâncias na avaliação genética de bovinos de corte. O autor utilizou, primeiramente, transformações logarítmicas e padronização pelo desvio padrão. Ao constatar a ineficiência das transformações, realizou análises por meio do modelo multicaracterístico, sendo o peso, em cada classe de desvio-padrão fenotípico, considerado como característica distinta. O autor concluiu que quando os touros são classificados por meio do modelo unicaracterístico, desconsiderando o efeito da heterogeneidade, há uma maior porcentagem de touros selecionados em comum entre os melhores indivíduos.

OLIVEIRA et al. (2001), trabalhando com dados reais de grupos genéticos formadores da raça Canchim, utilizaram o programa MTDFREML para calcular os componentes de variâncias e os valores genéticos das características peso aos 365 e 550 dias. A estratégia utilizada para estimar os componentes de variâncias para as duas características foi considerar a expressão das características em cada grupo genético como característica diferente. Dessa forma, foram testados diferentes modelos com intenção de verificar diferenças nas estimativas dos componentes de variâncias dos diferentes grupos genéticos. As estimativas dos componentes de variância e as classificações dos melhores animais foram diferentes, para o modelo unicaracter e o modelo tricaracter. Assim sendo, os autores concluíram que existe heterogeneidade de variâncias nos grupos genéticos participantes da formação da raça Canchim para as características analisadas.

CARVALHEIRO et al (2002), por meio de dados simulados, analisou o efeito da heterogeneidade de variâncias residuais, entre grupos de contemporâneo, sobre as avaliações genéticas. Comparou o uso de uma avaliação ponderada, por meio da flexibilização da matriz de (co)variância residual, em relação à avaliação que assume homogeneidade de variâncias. Foram simulados cinco diferentes níveis de heterogeneidade de variâncias residuais. O autor encontrou que em níveis mais acentuados de heterogeneidade de variâncias, os animais foram selecionados dos grupos mais variáveis, especialmente com pressão de seleção intensa e as avaliações genéticas ponderadas apresentaram resultados sempre iguais ou superiores

àqueles obtidos pelas avaliações que assumiram homogeneidade. Com relação à acurácia da predição, os produtos e as vacas tiveram seus valores genéticos preditos mais afetados pela heterogeneidade de variâncias residuais que os touros.

JENSEN e MADSEN (2002) apresentam um método Bayesiano para estimação de variâncias e covariâncias genéticas entre países e variâncias e covariâncias residuais dentro de países. O método incluiu a predição de valores genéticos, de tal forma que a incerteza sobre os componentes de variância seja considerada. Os autores utilizaram diferentes níveis de informação *a priori*, obtendo resultados similares às estimativas produzidas pelo REML quando *priors* não informativos foram empregados. Uma limitação do procedimento, ressaltada pelos autores, é que devem ser fornecidos iguais níveis de confiança para cada elemento da matriz de covariâncias. Entretanto, destacam que isto não constitui uma limitação prática, uma vez que a quantidade de informação *a priori*, de alguma forma, será sempre subjetiva. Portanto, é possível desenvolver um método com diferentes graus de confiança *a priori* para cada elemento.

CARNEIRO (2003), trabalhando com dados simulados, comparou análises que consideram ou não a presença de heterogeneidade. Verificou o efeito da heterogeneidade nas avaliações genéticas e seleção de touros, vacas e progênes, além de analisar a relação existente entre heterogeneidade e conexão genética dos dados. Neste estudo foram simulados vários conjuntos de dados, apresentando quatro estruturas de heterogeneidade: rebanhos com heterogeneidade para todos os parâmetros, tanto genéticos quanto fenotípicos; rebanhos com médias genéticas similares e demais parâmetros heterogêneos; rebanhos com heterogeneidade fenotípica e rebanhos sem heterogeneidade. Os rebanhos foram agrupados em três níveis: alta, média e baixa variabilidade. Foram simulados também diferentes graus de conexão entre os níveis de variabilidade. Foram realizadas análises com característica única e análises de características múltiplas, onde foi considerada a heterogeneidade. O autor concluiu, neste estudo, que o problema de heterogeneidade sobre as avaliações genéticas é devido, basicamente, à presença de médias genéticas diferentes entre rebanhos; as avaliações de vacas e progênes são mais afetadas pela heterogeneidade; o

modelo multicaracterístico não foi eficiente em eliminar os problemas de heterogeneidade sobre a avaliação genética; e o grau de conectividade dos dados influenciou os resultados das análises apenas quando os rebanhos tinham médias genéticas heterogêneas.

OLIVEIRA et al. (2004), trabalhando com dados reais de animais da raça Canchim, compararam a aplicação da análise Bayesiana na estimação de componentes de variância e covariância e na predição dos valores genéticos, com os resultados obtidos pela metodologia REML. Segundo os autores, a análise Bayesiana apresenta grande flexibilidade tanto em relação aos modelos que podem ser utilizados nas análises quanto às inferências que podem ser realizadas a partir dos resultados. Os autores concluíram que apesar dos resultados terem sido semelhantes, a metodologia Bayesiana é mais flexível permitindo inferências mais precisas.

FRIDRICH et al. (2004), trabalhando com dados simulados avaliaram o efeito da heterogeneidade de variâncias na classificação de aves, quando se utiliza o método frequentista REML – BLUP. Os autores concluíram que a heterogeneidade de variâncias causou grande mudança na ordem de classificação das aves, tanto para machos como para fêmeas.

BALIEIRO et al. (2002a) avaliaram diferentes tipos de transformação de dados, visando estabilizar a heterogeneidade de variâncias nos ganhos de peso da desmama ao sobreano em animais da raça nelore. Os dados foram transformados utilizando-se as funções de logaritmo na base dez, raiz quadrada, ponderação das observações subtraindo-se pela média da subclasse do grupo contemporâneo e dividindo-se pelo desvio padrão fenotípico da subclasse, e ponderação das observações dividindo-se pelo desvio padrão fenotípico da subclasse. Os autores encontraram que as transformações do ganho de peso em escala original e transformadas utilizando as funções de padronização pela média e desvio-padrão fenotípico da subclasse de grupo contemporâneo e a divisão pelo desvio-padrão fenotípico da subclasse de grupo contemporâneo, estabilizam as variâncias. Entretanto, verificaram uma tendência de redução nas relações entre os componentes de variância genética aditiva e residual, o que culminou em incrementos na magnitude da estimativa da herdabilidade.

BALIEIRO et al. (2002b), utilizando dados reais, verificaram o efeito da aplicação de um modelo multicaracterístico para considerar a presença de heterogeneidade de variâncias, verificaram ainda o seu impacto na avaliação genética de vacas e touros da raça nelore. Os autores utilizaram o método REML, por meio do programa MTDFREML para estimação dos componentes de variâncias e predição dos valores genéticos. Segundo os autores, com base nas estimativas das correlações genéticas entre pesos ao sobreano, correlações de Pearson e correlações de Spearman, desconsiderar a presença de heterogeneidade de variâncias não provocaria alterações profundas na classificação dos touros e das vacas para a característica ganho de peso da desmama ao sobreano.

Segundo TORRES (1998), a aplicação da metodologia BLUP é correta quando o modelo é apropriado e as matrizes de variâncias e covariâncias genéticas e residuais são conhecidas. Entretanto, na prática os componentes de variâncias nunca são conhecidos sem erros, o que se tem feito é, a partir dos dados disponíveis, estimá-los. Em populações com tamanho satisfatório sob normalidade e homogeneidade de variâncias, essa aproximação parece ser satisfatória. Porém, em situações com heterogeneidades de variância a dificuldade encontrada é em determinar quais componentes devem ser usados para cada ambiente.

Segundo BLASCO (2001), mesmo grandes conjuntos de dados com presença de heterogeneidade de variâncias, podem apresentar pouca informação. Nestas situações, métodos que necessitam de justificação assintótica, tal qual o método REML, podem produzir estimativas de componentes de variância não confiáveis dentro das subclasses.

Como visto, a heterogeneidade de variâncias representa um problema a ser resolvido, sendo que a solução mais comumente utilizada pela metodologia freqüentista, que é a utilização de um modelo misto multicaracterístico, pode resultar em um outro problema que é a estimação em pequenos conjuntos de dados.

Neste contexto, teoricamente a inferência Bayesiana apresenta-se como uma ferramenta promissora por levar em consideração a incerteza sobre os componentes de variâncias, permitir a inclusão de informação *a priori* e apresentar facilidade de implementação de distribuições mais robustas

2.2. A ESTATÍSTICA BAYESIANA

2.2.1. Teorema de Bayes

O teorema de Bayes foi formulado pelo reverendo Thomas Bayes no século XVII e caracteriza a metodologia estatística Bayesiana. O teorema de Bayes demonstra como as probabilidades mudam à luz de novas evidências. Antes de se analisar os dados, geralmente têm-se alguma informação prévia, se um dado é honesto ou não, se uma droga é efetiva, se um tratamento é eficiente e assim por diante. Esta informação *a priori* pode ser expressa em termos de probabilidades. Calcula-se dessa forma a verossimilhança atualizando sua probabilidade *a priori* usando o teorema de Bayes. Probabilidades *a priori* são baseadas em informações disponíveis separadamente do experimento e a verossimilhança deduzida do experimento. Dessa forma, é suposto que as probabilidades *a priori* representem o conhecimento antes que os dados sejam observados e que a verossimilhança expresse todo o conhecimento do “fenômeno” contido nos dados (LEANDRO, 2001).

Para o cálculo da probabilidade *a posteriori* de um modelo, multiplica-se sua probabilidade *a priori* pela sua verossimilhança, sendo este resultado dividido pela soma sobre todos os modelos para fazer a probabilidade total igual a um. Tem-se, então, que o teorema de Bayes pode ser utilizado para combinar a informação contida nos dados com a probabilidade *a priori*.

Considerando-se um vetor de parâmetros não observáveis θ que se deseja estimar, um vetor de observações y e a função densidade conjunta $f(\theta, y)$. Conforme a teoria da probabilidade:

$$f(\theta, y) = f(y|\theta) \cdot f(\theta) \quad [1]$$

e

$$f(\theta, y) = f(\theta) \cdot f(y|\theta) \quad [2]$$

onde $f(\cdot)$ e $f(\cdot)$ são as densidades marginais de \cdot e \cdot , respectivamente. Conforme [1] e [2], tem-se que:

$$f(\cdot) = f(\cdot) \cdot f(\cdot) / f(\cdot) \quad [3]$$

Para determinar a densidade marginal de \cdot , basta integrar a sua densidade conjunta em relação aos demais parâmetros. Assim, tem-se que:

$$f(\cdot) = \int_{\mathbb{R}} f(\cdot, \cdot) \cdot d = \int_{\mathbb{R}_0} f(\cdot) \cdot f(\cdot) d = E_{\cdot} [f(\cdot)], \quad [4]$$

onde E_{\cdot} representa a média com respeito à distribuição de \cdot . Nota-se, portanto, que $f(\cdot)$ não é uma função de \cdot . Logo, [3] pode ser escrita em forma de proporcionalidade:

$$f(\cdot) \propto f(\cdot) \cdot f(\cdot) \quad [5]$$

De acordo com a terminologia Bayesiana, $f(\cdot)$ é conhecida como a densidade *a priori* de \cdot , que reflete a incerteza relativa sobre os possíveis valores de \cdot antes das observações serem realizadas, ou seja, antes de se conhecer os valores do vetor \cdot . A densidade $f(\cdot)$ é a função de verossimilhança, que representa a contribuição de \cdot para o conhecimento de \cdot . Finalmente, $f(\cdot)$ é a densidade *a posteriori* que considera o estado de incerteza sobre \cdot dado um conhecimento prévio. Inferências sobre \cdot são feitas a partir da densidade *a posteriori* e a solução para o problema estatístico é finalizada com:

$$\text{Prob}(\cdot \in \mathbb{R}) = \int_{\mathbb{R}} f(\cdot) d, \quad [6]$$

onde \mathbb{R} é região do espaço de \cdot .

2.2.2. Os Métodos Bayesianos

Mais recentemente, inúmeras publicações vêm sendo feitas, demonstrando a utilização de métodos Bayesianos como uma poderosa ferramenta para resolução de problemas relacionados à avaliação genética dos animais.

No contexto Bayesiano, o objetivo é, dado as observações, descrever a incerteza sobre o real valor de algum parâmetro, usando a probabilidade como uma medida desta incerteza. Por exemplo, se o parâmetro de interesse é a herdabilidade de alguma característica, o objetivo da inferência Bayesiana é encontrar a densidade de probabilidade da herdabilidade dado os dados $f(h|\mathbf{y})$, em que \mathbf{y} é o vetor de observações. Quando esta distribuição é obtida, inferências podem ser feitas de múltiplas maneiras, como calcular a probabilidade da herdabilidade estar entre dois valores. Pode-se também determinar o menor intervalo no qual a probabilidade de encontrar a herdabilidade é maior que um determinado valor.

GIANOLA e FOULLEY (1982) introduziram os métodos Bayesianos no melhoramento animal no contexto de características de limiar, em seguida, GIANOLA e FERNANDO (1986) elucidaram novas possibilidades de utilização de técnicas Bayesianas. Entretanto, embora os métodos Bayesianos fossem teoricamente poderosos, usualmente eles conduziam a fórmulas na qual múltiplas integrais tinham que ser resolvidas para obtenção das distribuições marginais *a posteriori*, para uma completa inferência Bayesiana.

GIANOLA e FERNANDO (1986) apresentaram vantagens na utilização da inferência Bayesiana no melhoramento animal. Em associação com a teoria das probabilidades, mostraram que a inferência pode ser baseada na distribuição marginal *a posteriori* dos parâmetros de interesse, tal qual a incerteza sobre os demais parâmetros é completamente levada em consideração. O ponto de partida é a densidade conjunta *a posteriori* de todos os parâmetros desconhecidos. Por meio da distribuição conjunta, a distribuição marginal *a posteriori* de um parâmetro, como por exemplo o valor genético de um animal, é obtida sucessivamente integrando-se em ordem todos os parâmetros de “perturbação” (*nuisance*), sendo estes os efeitos sistemáticos, todos os efeitos aleatórios exceto aquele parâmetro de interesse, e os

componentes de variância e covariância. Esta integração é difícil ou praticamente impossível por meios analíticos ou numéricos, e assim normalmente buscam-se aproximações.

2.2.3. O Amostrador de Gibbs

Uma das maiores dificuldades técnicas na aplicação de métodos Bayesianos tem sido o processo de obtenção das distribuições marginais de interesse. Várias aproximações têm sido sugeridas e técnicas de integração numérica têm sido utilizadas. Segundo SORENSEN, (1986) a introdução de procedimentos *Markov Chain Monte Carlo* (MCMC) tem contribuído para a utilização de métodos Bayesianos no melhoramento animal. Dentre estes procedimentos, o Amostrador de Gibbs tem se demonstrado bastante útil e eficiente em problemas relacionados ao melhoramento animal, no qual envolvem vários modelos multidimensionais (GELFAND et al. 1990).

Assumindo-se um modelo que envolve um vetor de parâmetros $\theta = (\theta_1, \dots, \theta_n)$, o qual possui dois elementos, e o interesse é em θ_1 . A distribuição conjunta de θ é:

$$p(\theta_1, \theta_2, \dots, \theta_n) = p(\theta_1, \theta_2, \dots, \theta_n) \quad [7]$$

Inferências sobre θ_1 podem ser feitas a partir da distribuição marginal *a posteriori* de θ_1 .

$$p(\theta_1) = \int p(\theta_1, \theta_2, \dots, \theta_n) d\theta_2 \dots d\theta_n \quad [8]$$

Pode-se notar que a distribuição marginal *a posteriori* de θ_1 é a média das densidades condicionais de θ_1 , dado o parâmetro de “perturbação” $\theta_2, \dots, \theta_n$. O cálculo da média sobre os parâmetros de “perturbação” é o aspecto geral deste processo de marginalização. Por exemplo, θ_1 pode representar o valor genético, e $\theta_2, \dots, \theta_n$ os componentes de variância. Inferências sobre os valores genéticos são obtidas como uma média das densidades condicionais, com

igual importância para as densidades *a posteriori* dos componentes de variância.

A amostragem de Gibbs pode ser considerada como uma implementação prática do fato de que o conhecimento das distribuições condicionais é suficiente para determinar uma distribuição conjunta. As densidades condicionais completas *a posteriori* de todos os parâmetros desconhecidos, são necessárias para a realização da Amostragem de Gibbs. Estas distribuições correspondem à distribuição de um ou vários parâmetros especificados condicionais a todos os outros parâmetros do modelo. Em situações com um menor número de dimensões a derivação da marginal a partir das condicionais é direta. Entretanto, em um caso multidimensional, o problema torna-se mais complexo e estas conexões não são tão simples, sendo possível usar diferentes grupos de condicionais para calcular a marginal de interesse (WANG et al. 1994).

A amostragem de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, no qual são utilizadas como densidades candidatas distribuições completas condicionais.

Seja

$$p(\gamma_i | \gamma_j, j \neq i), \quad i = 1, \dots, k, \quad [9]$$

a distribuição condicional completa para todos os componentes individuais do vetor de quantidades desconhecidas γ , dado as observações. Supondo, então, que dado um conjunto arbitrário de pontos iniciais,

$$\gamma = (\gamma_1^0, \gamma_2^0, \dots, \gamma_k^0) \quad [10]$$

para a quantidade desconhecida, implementa-se o seguinte procedimento iterativo:

$$\text{gera-se } \gamma_1^1 \text{ de } p(\gamma_1 | \gamma_2^0, \gamma_3^0, \dots, \gamma_k^0);$$

$$\begin{aligned}
&\text{gera-se } \cdot_2^1 \text{ de } p(\cdot_2^1 | \cdot_1^0, \cdot_3^0, \dots, \cdot_k^0); \\
&\text{gera-se } \cdot_3^1 \text{ de } p(\cdot_3^1 | \cdot_2^1, \cdot_4^0, \dots, \cdot_k^0); \\
&\text{gera-se } \cdot_k^1 \text{ de } p(\cdot_k^1 | \cdot_1^1, \cdot_2^1, \dots, \cdot_{k-1}^1); \\
&\text{gera-se } \cdot_1^2 \text{ de } p(\cdot_1^2 | \cdot_2^1, \cdot_3^1, \dots, \cdot_k^1); \\
&\text{e assim por diante.}
\end{aligned}
\tag{111}$$

Supondo que o procedimento acima continue até a iteração t , a qual será repetida n vezes. Obtêm-se, dessa forma um vetor de amostras $\cdot^t = (\cdot_1^t, \dots, \cdot_k^t)$, onde \cdot^t é uma realização da cadeia de *Markov*. Quando $t \rightarrow \infty$, $\cdot^t = (\cdot_1^t, \dots, \cdot_k^t)$ tende em distribuição para a variável aleatória, cuja distribuição conjunta é dada por $p(\cdot)$. Em particular \cdot_i^t tende em distribuição à quantidade aleatória cuja densidade é $p(\cdot_i)$. Assim, para t grande, as réplicas $\cdot^t = (\cdot_1^t, \dots, \cdot_k^t)$ são aproximadamente amostras aleatórias da distribuição $p(\cdot)$ (GEMAN e GEMAN, 1984).

2.2.4. Distribuições *a priori*

Em muitas situações informações adicionais estão disponíveis sobre o valor do parâmetro desconhecido. Se esta informação pode ser utilizada para construir uma distribuição *a priori* para o parâmetro de interesse, então os métodos Bayesianos podem ser usados para estimar o desconhecido valor do parâmetro.

A distribuição *a priori* de um parâmetro é uma função de probabilidade ou função densidade de probabilidade expressando nosso grau de confiança sobre o valor do parâmetro, antes de se observar uma amostra da variável aleatória cuja função de distribuição depende do parâmetro (LARSON, 1982).

Existe uma grande discussão com relação à informação *a priori*, porque freqüentemente ela é subjetiva, representando o julgamento do indivíduo que está analisando os dados. Esta objeção está relacionada ao fato de que em alguns casos, principalmente em dados de pequeno tamanho, a distribuição *a posteriori* é muito sensível à escolha da *priori*, e assim, pesquisadores,

trabalhando com os mesmo dados, poderiam chegar a conclusões diferentes utilizando diferentes *priors*.

WANG et al. (1993) avaliaram o efeito de diferentes níveis de informação *a priori* em dados sob seleção, com o objetivo de ilustrar como informações externas podem influenciar os resultados. Neste estudo, quatro conjuntos de *priors* foram utilizados, sendo considerados mais informativos aqueles conjuntos com maiores quantidades de registros e medidas repetidas. Os autores observaram que as estimativas foram mais precisas quando o nível de informação *a priori* aumentou, demonstrando que a análise Bayesiana pode ser usada para incorporar informações externas em uma situação real de melhoramento animal.

2.2.5. Distribuições Robustas

Em genética quantitativa, a maioria das aplicações de modelos lineares mistos considera a distribuição Gaussiana tanto para os efeitos genéticos como para os efeitos residuais. Entretanto, esta suposição faz com que o modelo e seus resultados sejam vulneráveis à presença de valores discrepantes (Rogers e Tukey, 1972) citados por (ROSA, 2000). Uma das possibilidades que tem sido sugerida consiste em utilizar distribuições leptocúrticas ao invés da distribuição normal (SORENSEN e GIANOLA, 2000).

Segundo ROSA (2000), a aplicação de semelhantes distribuições no melhoramento animal é relativamente recente e tem sido concentrada na utilização da distribuição t. Entretanto, alternativas de distribuições para estimação robusta têm sido sugeridas, entre elas a distribuição *slash* e a normal contaminada. Estas distribuições pertencem a uma família denominada de distribuições normal/independentes.

A distribuição normal contaminada tem sido utilizada como uma opção eficiente para estimação robusta em análise de dados do melhoramento animal, sob diversas situações.

2.2.5.1. Distribuição Normal Contaminada

A vantagem da utilização de uma distribuição normal contaminada, em dados sob heterogeneidade, consiste basicamente na sua robustez. Estas distribuições são menos sensíveis ao desvio das pressuposições comumente

estabelecidas no melhoramento animal, tais como: normalidade dos dados; homogeneidade de variâncias; ausência de tratamento preferencial e assim por diante. No contexto de heterogeneidade de variâncias, estas distribuições apresentam-se menos sensíveis aos valores discrepantes, e conseqüentemente às possíveis diferenças que poderiam existir entre as variâncias de tratamentos. Obviamente, nem todas as situações de heterogeneidade de variâncias seriam resolvidas com a utilização destas distribuições. Entretanto, o emprego da distribuição normal contaminada é vantajoso nas situações onde a presença de valores contaminantes contribui para o aumento da heterogeneidade.

Seja $\mathbf{y} \sim N_{\lambda}(\boldsymbol{\mu}, \Sigma, \varphi, \lambda)$ um vetor aleatório de ordem k com distribuição normal contaminada com parâmetros $\boldsymbol{\mu}, \Sigma, \varphi$, e λ , o que corresponde a uma mistura de duas distribuições normais $N(\boldsymbol{\mu}, \lambda^{-1}\Sigma)$ e $N(\boldsymbol{\mu}, \Sigma)$, nas proporções de φ e $(1-\varphi)$, respectivamente (LITTLE, 1988). A densidade desta distribuição é dada por:

$$p(\mathbf{y}; \boldsymbol{\mu}, \Sigma, \varphi, \lambda) \propto |\Sigma|^{-1/2} \left[\varphi \lambda^{k/2} \exp\left\{-\frac{\lambda}{2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})\right\} + (1-\varphi) \exp\left\{-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})\right\} \right] \quad [12]$$

com $-\infty < y_i < \infty$, $-\infty < \mu_i < \infty$, $|\Sigma| > 0$, $0 \leq \varphi < 1$, e $0 \leq \lambda < 1$

em que $\mathbf{y} = (y_1, \dots, y_k)$, $\boldsymbol{\mu}$ é a proporção de valores discrepantes, e λ é o fator que exprime a relação entre variâncias e covariâncias das distribuições referentes às populações contaminante e não-contaminante.

A distribuição de \mathbf{y} é simétrica e unimodal, média $\boldsymbol{\mu}$ e matriz de covariâncias $(1-\varphi + \varphi/\lambda)\Sigma$. Algumas propriedades da distribuição normal multivariada também são asseguradas para variáveis com distribuição normal contaminada multivariada, por exemplo, distribuições marginais condicionais, bem como qualquer combinação linear das mesmas, também possuem distribuição normal contaminada. Uma importante exceção se refere à independência e correlação. Elementos de uma distribuição normal multivariada com matriz de covariâncias diagonal são não correlacionados e

independentes. Entretanto, numa situação deste tipo com a distribuição normal contaminada, têm-se elementos não correlacionados, mas não independentes (ROSA 2000).

A distribuição normal contaminada pode ser obtida como a distribuição marginal de y em relação a w , sendo $y \sim N(\mu, \Sigma + w^{-1}\Sigma)$, onde w é uma variável discreta concentrada em dois pontos, λ e 1 (LITTLE, 1988), com distribuição de probabilidades dada por:

$$P(W|\varphi, \lambda) = \begin{cases} \varphi & , \text{ se } w = \lambda \\ 1 - \varphi & , \text{ se } w = 1 \end{cases} \quad [13]$$

Com $0 \leq \varphi < 1$, e $0 \leq \lambda < 1$.

As densidades *a priori*, as densidades condicionais completas e a densidade conjunta *a posteriori* da distribuição normal contaminada para os resíduos são dadas com maiores detalhes em ROSA (1999).

PEREIRA et al. (2002), utilizando dados simulados, compararam estimativas de componentes de variâncias e predição de valores genéticos, obtidos por meio de um modelo linear misto, na qual foi utilizada distribuição Gaussiana e distribuição normal contaminada para os resíduos, via amostrador de Gibbs. O processo de simulação consistiu de 50 arquivos com 1000 indivíduos cada, seguido de 5 gerações, com dois níveis de efeitos fixos, e características onde foram introduzidos diferentes níveis de contaminação. Os autores encontraram que, exceto para a população sem contaminantes, o modelo com distribuição normal contaminada apresentou melhores estimativas de variâncias residuais e os valores genéticos preditos se aproximaram dos valores reais. Dessa forma, os autores concluíram que o modelo linear misto com distribuição normal contaminada para os resíduos, apresenta-se como uma alternativa para estimação robusta no melhoramento animal.

SANTOS et al. (2004) trabalharam com três modelos para predição de valores genéticos para bovinos da raça Simental: modelo gaussiano implementado pelo método freqüentista REML; modelo gaussiano implementado sob análise Bayesiana; e modelo robusto aplicado sob a pressuposição de distribuição normal contaminada para os dados. Segundo os

autores, uma alternativa viável para corrigir o problema de heterogeneidade de variâncias é a utilização das distribuições normal-independentes que podem ser aplicadas por meio de modelos Bayesianos. Os autores concluíram que a utilização do modelo robusto para estimação dos componentes de variância e predição dos valores genéticos de animais da raça Simental pode levar a um aumento na acurácia da predição, para peso aos 730 dias e, conseqüentemente, a uma maior resposta à seleção quando há presença de heterogeneidade de variâncias nos dados.

3. MATERIAL E MÉTODOS

A simulação computacional tem demonstrado ser uma poderosa ferramenta no auxílio à implementação e emprego adequado de metodologias que visam solucionar problemas relacionados ao melhoramento animal, entre eles, métodos de seleção, testes de pressuposições, verificação de novas metodologias, conexão dos dados, qualidade dos dados, heterocedasticidade, entre outros.

Os dados utilizados neste trabalho foram simulados utilizando-se o sistema computacional Genesys (EUCLYDES, 1996). O sistema é escrito na linguagem de programação FORTRAN, permite a criação de genomas de certa complexidade, que podem ser utilizados para formação de populações, com estruturas desejadas, sob a influência dos problemas propostos a serem analisados. A possibilidade da simulação de populações, desde a criação do genoma, introdução de efeitos ambientais, respeitando-se todas as pressuposições estatísticas estabelecidas em populações reais, permite ao sistema Genesys simular dados consistentes para verificação em análises.

3.1. SIMULAÇÃO DO GENOMA

Foi simulado um genoma de 3000 centimorgans de comprimento, considerando uma única característica quantitativa, governada por 800 locos com dois alelos por loco, na qual a herdabilidade variou conforme a estrutura desejada de heterogeneidade de variâncias.

Os locos quantitativos foram simulados segundo a distribuição normal e estavam distribuídos em vinte pares de cromossomos de tamanhos aleatórios. Os locos não possuíam desvios de dominância nem interações entre si. O

genoma não possuía cromossomo sexual e as frequências alélicas iniciais eram iguais em ambos os sexos.

A distribuição matemática usada para a simulação das frequências iniciais dos locos quantitativos nas populações foi a uniforme, com média de 0,50. Foi utilizado um efeito fixo de rebanho com 6 níveis e os efeitos de ambiente não-controláveis foram simulados conforme a distribuição normal. O processo de simulação do genoma teve 100 repetições.

3.2. SIMULAÇÃO DAS POPULAÇÕES

Segundo a estrutura genômica proposta, foram simulados 1500 machos e 1500 fêmeas que formaram a população base.

A partir da população-base foram formadas duas populações iniciais, uma grande e outra pequena. Para a formação da população inicial grande foram escolhidos aleatoriamente 50 machos e 1500 fêmeas, na razão de um macho para 30 fêmeas, que deram origem a 3000 descendentes (dois descendentes por fêmea). Esta população foi submetida a uma geração de seleção ao acaso para a formação da população inicial grande com 6000 descendentes. Para formação da população inicial pequena foram escolhidos aleatoriamente 4 machos e 120 fêmeas, na razão de um macho para 30 fêmeas, que deram origem a 240 descendentes (dois descendentes por fêmea). Esta população foi submetida a uma geração de seleção ao acaso para a formação da população inicial pequena com 480 descendentes.

Com o propósito de se avaliar o efeito dos diferentes tipos de heterogeneidade de variâncias, com diferentes tamanhos de população, bem como comparar os métodos frequentista REML e BLUP com um método Bayesiano de análise, foram inseridos diferentes tipos de estruturas de heterogeneidade nas populações iniciais. Para obtenção destas estruturas de heterogeneidade foram feitos descartes estratégicos dos valores máximos e mínimos genéticos, ambientais, ou de ambos, de acordo com o tipo de heterogeneidade e o nível de variabilidade desejada. Nos Quadros 1 e 2 são apresentados a forma de obtenção das populações finais grande e pequena e os códigos utilizados para representar as populações com diferentes tipos de

heterogeneidade de variâncias, níveis de variabilidade e tamanhos de população, respectivamente.

Quadro 1 – Número de indivíduos das populações iniciais grande e pequena, subpopulações iniciais e finais grande e pequena e das populações finais.

Populações Iniciais	Subpopulações Iniciais	Subpopulações Finais	Populações Finais
Grande (6000)	2000	1600	Grande (4800)
	2000	1600	
	2000	1600	
Pequena (480)	160	100	Pequena (300)
	160	100	
	160	100	

Quadro 2 - Código utilizado para as populações com diferentes tipos de heterogeneidade de variâncias, níveis de variabilidade e tamanhos de população.

Estrutura dos Dados	Nível de Variabilidade	Tamanho da População	
		População Pequena (P)	População Grande (G)
Homogeneidade de Variâncias (HO)	População 1 (1)	HO-P-1	HO-G-1
	População 2 (2)	HO-P-2	HO-G-2
	População 3 (3)	HO-P-3	HO-G-3
	Total (T)	HO-P-T	HO-G-T
Heterogeneidade de Variância Genética (HG)	Alta (A)	HG-P-A	HG-G-A
	Média (M)	HG-P-M	HG-G-M
	Baixa (B)	HG-P-B	HG-G-B
	Total (T)	HG-P-T	HG-G-T
Heterogeneidade de Variância Ambiental (HA)	Alta (A)	HA-P-A	HA-G-A
	Média (M)	HA-P-M	HA-G-M
	Baixa (B)	HA-P-B	HA-G-B
	Total (T)	HA-P-T	HA-G-T
Heterogeneidade de Variância Genética e Ambiental (HGA)	Alta (A)	HGA-P-A	HGA-G-A
	Média (M)	HGA-P-M	HGA-G-M
	Baixa (B)	HGA-P-B	HGA-G-B
	Total (T)	HGA-P-T	HGA-G-T

$$T = (A+M+B) \text{ ou } T = (\text{População1} + \text{População2} + \text{População3})$$

Em todos os níveis de heterogeneidade estudados, as populações possuíam boa estrutura de conectividade alcançada pela utilização de touros e vacas utilizados em comum entre as populações.

Nas populações com heterogeneidade de variância genética aditiva foram feitos descartes dos valores genéticos reais máximos e mínimos. Nas populações com heterogeneidade de variância ambiental foram feitos descartes nos valores ambientais reais máximos e mínimos. Nas populações com heterogeneidade de variância genética aditiva e ambiental foram feitos descartes nos valores genéticos e ambientais reais máximos e mínimos.

3.3. ANÁLISE DOS DADOS

Os dados apresentando diferentes tipos de heterogeneidade, com diferentes tamanhos de população e diferentes níveis de variabilidade, foram submetidos à análise para estimação dos componentes de variância e predição dos valores genéticos, por meio de duas metodologias: a combinação dos métodos freqüentistas REML – BLUP (Máxima Verossimilhança Restrita – Melhor preditor Linear não viesado, via equações de modelos mistos de Henderson) e o método Bayesiano, o qual combina informações *a priori* e *a posteriori*.

Para ambas as metodologias, foram realizados dois conjuntos de análises, sendo que para a metodologia Bayesiana foi utilizada também uma distribuição normal contaminada para os resíduos. Esta forma de análise foi empregada para a população com homogeneidade de variâncias e para a população com heterogeneidade de variância ambiental.

No primeiro conjunto de análises, desconsiderou-se a presença de heterogeneidade de variâncias admitindo-se que todas as subpopulações pertenciam à mesma população, sendo as análises realizadas com a população total.

No segundo conjunto, as análises foram realizadas isoladamente para cada subnível de variabilidade, como descrito no Quadro 2.

O seguinte modelo linear misto unicaracterístico foi empregado nos dois conjuntos de análises:

$$y_{ij} = \mu + R_i + a_{ij} + e_{ij}, \quad [14]$$

em que y_{ij} é a observação fenotípica do $j^{\text{ésimo}}$ animal no $i^{\text{ésimo}}$ rebanho; μ é uma constante inerente a todas as observações; R_i é o efeito do $i^{\text{ésimo}}$ rebanho;

a_{ij} efeito aleatório genético aditivo do $j^{\text{ésimo}}$ animal no $i^{\text{ésimo}}$ rebanho R; e_{ij} é o efeito aleatório ambiental.

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad [15]$$

em que,

\mathbf{y} = vetor das observações dos indivíduos;

\mathbf{X} = matriz de incidência de efeitos *fixos*;

$\boldsymbol{\mu}$ = vetor de efeitos *fixos* (rebanho);

\mathbf{Z} = matriz de incidência dos efeitos aleatórios;

\mathbf{a} = vetor de efeitos aleatórios (valores genéticos dos indivíduos);

\mathbf{e} = vetor de erros aleatórios.

As estimativas dos componentes de variâncias e dos parâmetros genéticos pelo método freqüentista foram obtidas utilizando-se o programa MTDFREML (Multiple Trait Derivative-Free Restricted Maximum Likelihood), descrito por BOLDMAN et al. (1995), empregando-se a Metodologia de Modelo Mistos de Henderson. Para o método Bayesiano, as estimativas foram obtidas utilizando-se o programa MTGSAM (Multiple Trait Gibbs Sampling in Animal Models), descrito por VAN TASSEL e VAN VLECK (1995). Para as análises realizadas, pela metodologia Bayesiana, foram utilizados três níveis de informações *a priori*: *priors* não informativos (NI), *priors* pouco informativos (PI) e *priors* informativos (I). Estes níveis de informação foram obtidos variando-se os hiperparâmetros da distribuição *a priori* utilizada para os componentes de variância (distribuição de qui-quadrado escalonada invertida).

3.3.1. Pressuposições e Distribuições a priori

A distribuição *a priori* utilizada para os componentes de variância foi a quiquadrado escalonada invertida, com a seguinte função densidade de probabilidade:

$$p(\sigma_i^2 | v_i, s_i^2) \propto (\sigma_i^2)^{-\left(\frac{v_i}{2} + 1\right)} \exp\left(-\frac{v_i s_i^2}{2\sigma_i^2}\right) \quad (i = a, e), \quad [16]$$

em que, v_i é o grau de confiança do componente de variância i , s_i^2 pode ser interpretado como o valor *a priori* para os componentes de variância e s_i^2 refere-se à distribuição marginal do componente de variância i , dado os hiperparâmetros. De acordo com os parâmetros dessa distribuição (v, s^2) , foram obtidos três conjuntos de *priors* com diferentes níveis de informação a respeito dos componentes de variância. No primeiro conjunto de *priors*, foram assumidos valores de $v = -2$ e $s = 0$, que indicou ausência de conhecimento sobre os componentes de variância. No segundo conjunto de *priors*, foram assumidos valores de $v = 5$ para a variância genética aditiva, $v = 15$ para a variância ambiental e s igual ao valor real do componente de variância a ser estimado, o que indicou *priors* pouco informativos sobre os componentes de variância. No terceiro conjunto de *priors*, foram assumidos valores de $v = 60$ para a variância genética aditiva, $v = 100$ para a variância ambiental e s igual ao valor real do componente de variância a ser estimado, o que indicou *priors* informativos sobre os componentes de variância a ser estimado. Assim, foram considerados diferentes níveis de informação *a priori*, desde *priors* não-informativos até *priors* próximos do real valor do parâmetro (informativos).

Para o vetor de observações \mathbf{y} , pressupõe-se que segue distribuição normal multivariada:

$$\mathbf{y} | \boldsymbol{\mu}, s_a^2, s_e^2 \sim N(\mathbf{X}\boldsymbol{\mu} + \mathbf{Z}\mathbf{a}, \mathbf{I} s_e^2) \quad [17]$$

A distribuição *a priori* do vetor $\boldsymbol{\mu}$ é representada por uma constante, indicando que se possui pouco conhecimento sobre os valores de seus elementos:

$$f(\boldsymbol{\mu}) \propto \text{constante} \quad [18]$$

Para o valor genético, considerou-se uma distribuição normal, de forma que:

$$\mathbf{a} | s_a^2 \sim N(\mathbf{0}, \mathbf{A} s_a^2) \quad [19]$$

em que A_i é uma matriz conhecida do numerador dos coeficientes de parentescos entre os indivíduos e s_a^2 é a variância da distribuição *a priori* dos valores genéticos \mathbf{a} . Todos os valores genéticos foram admitidos *a priori* como mutuamente independentes, assim como independentes também dos efeitos *fixos*.

Para o vetor de resíduos considerou-se distribuição normal ou distribuição normal contaminada:

$$e_i \sim \mathcal{N}(0, s_e^2), \quad \text{para distribuição Gaussiana.} \quad [20]$$

$$e_i \stackrel{C}{\sim} \mathcal{N}(0, \sigma^2, \phi, \lambda), \quad \text{para distribuição Normal Contaminada.} \quad [21]$$

O modelo com distribuição normal contaminada para os resíduos possui uma mistura de escalas da distribuição normal usada no lugar da usual suposição Gaussiana. Neste caso, cada elemento de $\boldsymbol{\varepsilon}_i$ é definido como $\omega_i = e_i / \sqrt{w_i}$, $i = 1, 2, \dots, n$, onde w_i é uma variável aleatória com densidade

$p(w_i|v)$, independente de e_i , onde $e_i \sim \mathcal{N}(\mu, \sigma^2)$, v é um parâmetro de robustez e ν é o fator que exprime a relação entre variâncias e covariâncias referentes às populações contaminantes e não contaminantes (ROSA, 1999).

3.3.2. Critério de Convergência

Considerando que a convergência foi atingida após um determinado número de iterações, os valores gerados nos primeiros ciclos foram descartados, o período de “burn in” utilizado foi de 100.000 iterações para populações pequenas e 10.000 iterações para populações grandes e o tamanho da cadeia foi de 1.000.000 e 200.000 amostras para as populações pequenas e grandes, respectivamente. O intervalo de retirada de amostras foi de 20 e 50 para populações grande e pequena, respectivamente.

O critério de convergência adotado foi o da inspeção visual dos gráficos e o programa GIBANAL (VANKAAN, 1998).

3.4. CRITÉRIOS DE AVALIAÇÃO E VARIÁVEIS AVALIADAS

3.4.1. Acurácia das estimativas das variâncias

Para avaliação dos valores, dos componentes de variâncias, estimados pelas metodologias REML e Bayesiana, em todas as estruturas de heterogeneidade e em todos os níveis de variabilidades estudados, calculou-se a porcentagem de erro, por meio da seguinte expressão:

$$PE = \frac{(\hat{\sigma}_i^2 - \sigma_i^2)100}{\sigma_i^2}, \quad i = a, e \quad [22]$$

em que PE é a porcentagem de erro na estimação do componente de variância; $\hat{\sigma}_i^2$ é o componente de variância estimado, σ_i^2 é o componente de variância real; $i=a,e$ refere-se às variâncias genéticas aditivas e ambientais, respectivamente.

Por meio da porcentagem de erro na estimação dos componentes de variância é possível verificar qual o efeito do tipo de heterogeneidade de

variâncias na estimação dos componentes de variâncias pelos métodos estudados.

Quanto mais próximos os valores estiverem de zero mais acurada é a estimativa.

3.4.2. Correlações entre os valores genéticos

Após a predição dos valores genéticos pelos métodos freqüentistas e Bayesiano, calculou-se a correlação de ordem entre valores genéticos preditos e reais, que eram conhecidos, uma vez que foram simulados pelo programa Genesys.

As correlações de ordem entre valores genéticos preditos e reais foram obtidas pelo SAS - *Statistical Analysis System*, (LITTEL et al. 1991). Por meio da correlação de Spearman foi possível verificar o efeito do tipo de heterogeneidade, do nível de variabilidade, do método estudado e do nível de informação *a priori* na classificação dos indivíduos.

Para verificar o efeito dos diferentes tipos de heterogeneidade e tamanhos de população sobre a seleção, comparou-se a porcentagem de indivíduos em comum entre os 15% melhores indivíduos de acordo com seus valores genéticos reais e preditos, obtidos pelas metodologias em análise. Procedeu-se também o cálculo da porcentagem de indivíduos selecionados por ambiente entre os 15% melhores valores genéticos preditos.

3.4.3. Quadrado Médio do erro (QME)

Os valores genéticos preditos foram comparados com os valores genéticos reais, por meio da média do erro ao quadrado, como em Harvey (1979), citado por SCHENKEL (1998).

$$QME = \sum_{i=1}^n \frac{1}{n} (\hat{a}_i - a_i)^2 \quad [23]$$

em que QME é o quadrado médio do erro, n é o número de indivíduos, \hat{a}_i e a_i são os valores genéticos predito e real do i ^{ésimo} indivíduo, respectivamente.

Quanto mais próximo de zero maior é a acurácia da predição dos valores genéticos.

3.4.4. Cálculo dos Componentes de Variâncias Reais

Os componentes de variância genética aditiva e ambiental reais, foram calculados segundo a fórmula:

$$\sigma_i^2 = \frac{\sum_{j=1}^n (w_i)_j^2 - \left[\sum_{j=1}^n (w_i)_j / n \right]^2}{n-1} \quad (i=a, e), \quad [24]$$

em que w_a e w_e são os valores genético-aditivos e ambientais reais respectivamente, e n é o número de observações.

4. RESULTADOS E DISCUSSÃO

4.1. VALORES REAIS DOS COMPONENTES DE VARIÂNCIA

Os valores reais dos componentes de variâncias genéticas aditivas e ambientais, valores de herdabilidades e número de indivíduos simulados para as populações com homogeneidade de variância, heterogeneidade de variância genética aditiva, heterogeneidade de variância ambiental e heterogeneidade de variância genética aditiva e ambiental, com diferentes níveis de variabilidade e dois tamanhos, são apresentados nos Quadros 3 a 6. O comportamento dos parâmetros simulados está próximo de uma situação real para a característica peso ao sobreano, em rebanhos da raça nelore, como descrito por BALIEIRO (2001). Quando a presença de heterogeneidade foi desconsiderada, as análises foram conduzidas com as populações totais, compostas de 300 indivíduos para a população pequena e 4800 indivíduos para a população grande.

Quadro 3 - Valores reais dos componentes de variância genética aditiva e ambiental e de herdabilidade das populações pequenas e grandes com iguais níveis de variabilidade genética aditiva e ambiental.

Populações	σ_a^2	σ_e^2	h^2	Nº de Indivíduos
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	1600
HO-G-2	309,22	236,46	0,57	1600
HO-G-3	318,21	240,93	0,57	1600
HO-G-T	310,55	245,04	0,56	4800

Quadro 4 - Valores reais dos componentes de variância genética aditiva e ambiental e de herdabilidade das populações pequenas e grandes com três níveis de variabilidade genética aditiva.

Populações	σ_a^2	σ_e^2	h^2	Nº de Indivíduos
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	1600
HG-G-M	191,77	237,15	0,45	1600
HG-G-B	140,46	252,15	0,36	1600
HG-G-T	212,05	249,01	0,46	4800

Quadro 5 - Valores reais dos componentes de variância genética aditiva e ambiental e de herdabilidade das populações pequenas e grandes com três níveis de variabilidade ambiental.

Populações	σ_a^2	σ_e^2	h^2	Nº de Indivíduos
HA-P-A	268,14	1719,20	0,13	100
HA-P-M	257,34	828,11	0,24	100
HA-P-B	274,45	411,80	0,40	100
HA-P-T	266,65	1009,54	0,21	300
HA-G-A	271,35	1372,21	0,17	1600
HA-G-M	273,62	960,50	0,22	1600
HA-G-B	249,84	628,04	0,29	1600
HA-G-T	265,04	987,43	0,21	4800

Quadro 6 - Valores reais dos componentes de variância genética aditiva e ambiental e de herdabilidade das populações pequenas e grandes com três níveis de variabilidade genética aditiva e ambiental.

Populações	σ_a^2	σ_e^2	h^2	Nº de Indivíduos
HGA-P-A	259,29	1302,90	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	1414,14	0,19	1600
HGA-G-M	194,26	868,05	0,18	1600
HGA-G-B	123,56	476,68	0,21	1600
HGA-G-T	219,47	919,85	0,19	4800

Para a estrutura de dados com heterogeneidade genética aditiva (Quadro 4), a herdabilidade foi variável nas subpopulações, sendo maior no subnível de maior variabilidade e menor no subnível de menor variabilidade. Existem evidências de trabalhos anteriores (HILL et al. 1983, BROTHERSTONE e HILL, 1986) de que a herdabilidade tende ser em média maior em ambientes mais variáveis. Contudo, na prática, a maior dificuldade é identificar qual rebanho apresenta maior variabilidade. Para a estrutura com heterogeneidade ambiental (Quadro 5), ocorreu o inverso, o ambiente com maior variabilidade ambiental apresentou menor herdabilidade e o ambiente com menor variabilidade apresentou a maior herdabilidade. Este fato ocorreu porque, no processo de simulação da heterogeneidade para estas estruturas, variou-se apenas um dos componentes, de acordo com a estrutura desejada. Exceção feita para a estrutura com heterogeneidade genética aditiva e ambiental em que os dois componentes variaram na mesma direção, ou seja, rebanhos com maior variabilidade ambiental apresentaram também maior variabilidade genética aditiva e vice-versa. Nesta situação as herdabilidades apresentaram-se mais homogêneas entre os ambientes (Quadro 6). Outras situações de heterogeneidade de variância genética aditiva e ambiental podem ocorrer na prática, entretanto, não foram abordadas neste trabalho.

Para a estrutura de população com homogeneidade de variâncias, utilizada como população controle, desejou-se que todos os componentes fossem similares e que as três subpopulações formadas apresentassem herdabilidade próximas entre si (Quadro 3).

4.2. ESTIMAÇÃO DOS COMPONENTES DE VARIÂNCIA

4.2.1. Análises pela Metodologia REML

4.2.1.1. Análises desconsiderando a Heterogeneidade

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

No Quadro 7 são apresentados as estimativas dos componentes de variância genética aditiva e ambiental e de herdabilidade, obtidos pela metodologia REML, para as populações totais pequenas e grandes com homogeneidade de variâncias, heterogeneidade de variância genética, heterogeneidade de variância ambiental e heterogeneidade de variância genética e ambiental. Pôde-se verificar que nas populações pequenas, em todas as estruturas analisadas, os valores de variância genética aditiva foram subestimados com relação ao valor real da população total. A estrutura de população com heterogeneidade genética e ambiental apresentou as piores estimativas, apresentando valor de herdabilidade igual a 0,02, demonstrando que a heterogeneidade nos dois componentes pode causar maiores problemas nas estimativas, quando a heterogeneidade é desconsiderada. Para a população com homogeneidade de variâncias e tamanho grande as estimativas obtidas ficaram mais próximas da situação real. Em análises prévias, foi verificado que as subpopulações de tamanho pequeno apresentavam maiores problemas devido à normalidade dos dados.

Para verificação do efeito de se desconsiderar a heterogeneidade de variâncias, as estimativas dos componentes de variâncias, obtidas a partir da população total, foram comparadas com os valores reais dos componentes do nível de interesse: alta, média ou baixa variabilidade.

Quadro 7 - Valores dos componentes de variância genética aditiva e ambiental e de herdabilidade das populações totais pequenas e grandes, com homogeneidade de variâncias e presença de heterogeneidade de variância genética aditiva, heterogeneidade de variância ambiental e heterogeneidade de variância genética aditiva e ambiental, estimados pela metodologia REML.

Populações	\hat{S}_a^2	\hat{S}_e^2	\hat{h}^2	Nº de Indivíduos
HO-P-T	64,29	366,35	0,15	300
HO-G-T	370,73	209,83	0,64	4800
HG-P-T	44,73	363,26	0,11	300
HG-G-T	201,79	274,57	0,42	4800
HA-P-T	323,43	841,40	0,28	300
HA-G-T	235,04	1021,89	0,19	4800
HGA-P-T	17,94	1053,30	0,02	300
HGA-G-T	160,81	998,26	0,14	4800

4.2.1.2. Porcentagens de Erros

Nos gráficos das Figuras de 1 a 4 são apresentados as porcentagens de erros das estimativas obtidas pelo método REML, ao calcular os componentes de variância genética aditiva e ambiental, desconsiderando a presença de heterogeneidade. Os resultados são expressos em porcentagem de erro em relação aos valores reais dos componentes de variância.

Quando se desconsiderou a presença de heterogeneidade de variâncias as análises foram realizadas com a população total de 4800 indivíduos para o tamanho grande e com a população total de 300 indivíduos para a população de tamanho pequeno. Pode-se observar que a maioria das estimativas obtidas, para as populações pequenas, foi inconsistente, indicando que a subpopulação com maior problema de normalidade dos dados pode estar afetando todo o conjunto.

Com relação ao tipo de heterogeneidade, observou-se que a heterogeneidade genética aditiva afetou a estimação da variância genética aditiva, a heterogeneidade ambiental afetou a estimação da variância ambiental e a heterogeneidade genética aditiva e ambiental afetou ambas as variâncias genética aditiva e ambiental. Este fato foi potencializado para as populações pequenas.

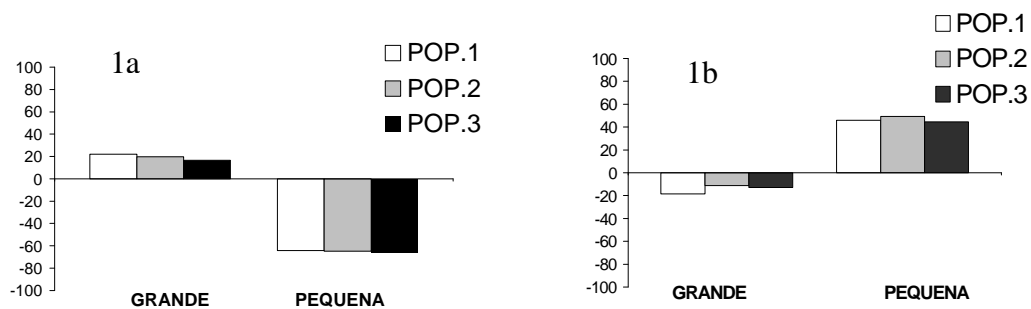


Figura 1. Percentagens de erro das estimativas dos componentes de variância genética aditiva (1a) e ambiental (1b), obtidas pela metodologia REML, em populações com 300 indivíduos (PEQUENA) e populações com 4800 indivíduos (GRANDE), sob homogeneidade de variâncias e iguais níveis de variabilidade.

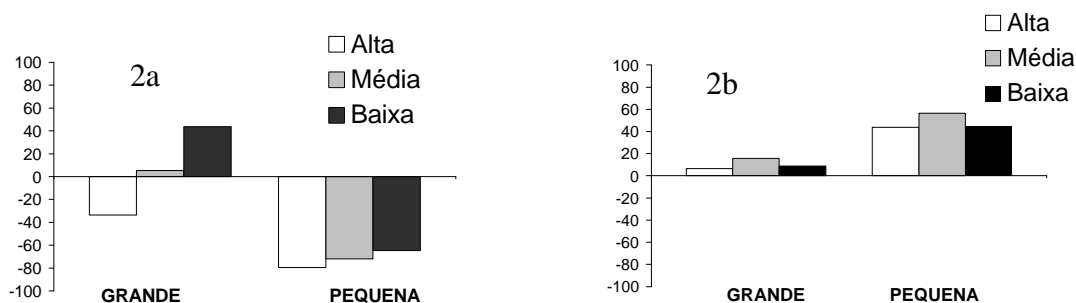


Figura 2. Percentagens de erro das estimativas dos componentes de variância genética aditiva (2a) e ambiental (2b), obtidos pela metodologia REML, em populações com 300 indivíduos (PEQUENA) e populações com 4800 indivíduos (GRANDE), sob heterogeneidade de variância genética aditiva em três níveis de variabilidade.

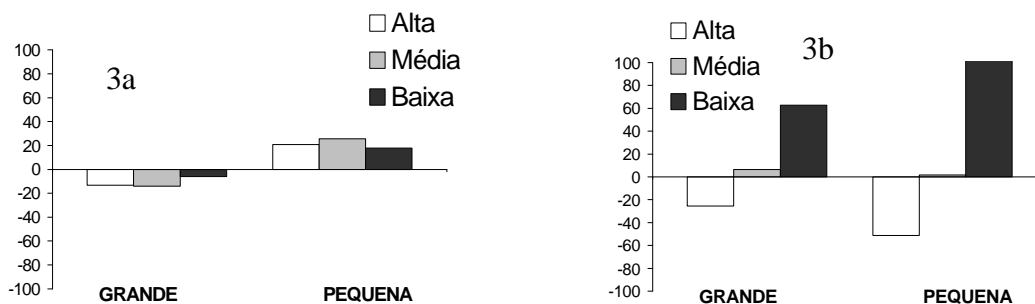


Figura 3. Percentagens de erro das estimativas dos componentes de variância genética aditiva (3a) e ambiental (3b), obtidos pela metodologia REML, em populações com 300 indivíduos (PEQUENA) e populações com 4800 indivíduos (GRANDE), sob heterogeneidade de variância ambiental em três níveis de variabilidade.

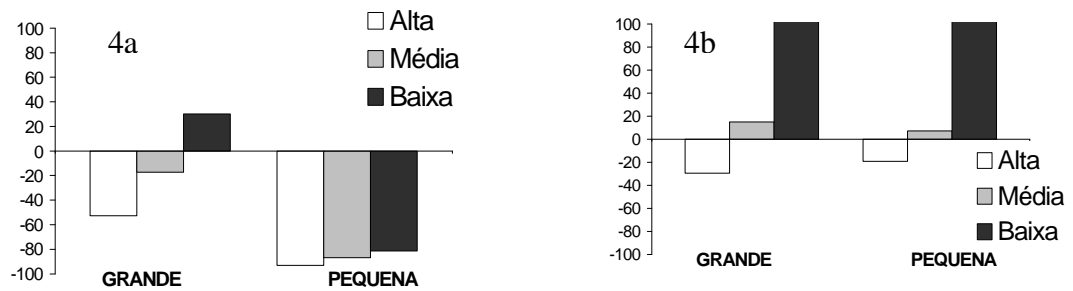


Figura 4. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (4a) e ambiental (4b), obtidos pela metodologia REML, em populações com 300 indivíduos (PEQUENA) e populações com 4800 indivíduos (GRANDE), sob heterogeneidade de variância genética aditiva e ambiental em três níveis de variabilidade.

As populações com níveis de variabilidade média foram, em algumas situações, menos afetadas pela presença de heterogeneidade, por possuir valores de variâncias próximos da média das variâncias das três populações (FIGURA 2a, 3b, 4a e 4b).

Para as populações grandes, apesar de não terem ocorrido problemas na convergência e as estimativas terem sido mais acuradas, quando a heterogeneidade não foi considerada alguns níveis de variabilidade tiveram seus valores subestimados ou superestimados, como pode ser observado no gráfico 2a, onde as populações com níveis de variabilidade alta e baixa tiveram suas estimativas de variância genética aditiva subestimadas e superestimadas, respectivamente. No gráfico 3b, onde a heterogeneidade ambiental foi desconsiderada, é possível verificar que para a variância ambiental também ocorreram valores subestimados e superestimados para as subpopulações com níveis de variabilidade alta e baixa. Contudo, as estimativas obtidas para o componente de variância homogêneo, para as subpopulações de tamanho grande, foram sempre acuradas e regulares nos três níveis de variabilidade, como pode ser visto no gráfico 2b e 3a. Por meio destes resultados pode-se concluir que a heterogeneidade de variância para um dos componentes não afeta a estimação do outro quando o tamanho da população é grande.

A partir da verificação do problema de se desconsiderar a presença de heterogeneidade de variâncias, procederam-se as análises, utilizando-se um modelo multicaracterístico, em que cada ambiente foi considerado como uma característica distinta. Entretanto, o emprego deste modelo não se mostrou

adequado, para as estruturas de dados utilizadas neste estudo, não ocorrendo convergência.

4.2.1.3. Análises dentro dos Subníveis

Como o modelo multicaracterístico não se mostrou adequado, optou-se pelas análises de característica única, considerando cada subpopulação isoladamente. As estimativas dos componentes de variância genética aditiva, variância ambiental e de herdabilidades, obtidas pela metodologia REML, nas populações de tamanho grande e tamanho pequeno, considerando cada ambiente isoladamente, são apresentados nos Quadros 8 a 11.

Para as subpopulações oriundas da população grande, com homogeneidade de variâncias (Quadro 8), as estimativas de herdabilidade de 0,47, 0,53 e 0,65, para as subpopulações 1, 2, e 3, respectivamente, ficaram próximas das herdabilidades simuladas de 0,54, 0,57 e 0,57. Entretanto, para as demais subpopulações oriundas das populações com heterogeneidade de variâncias, as estimativas de herdabilidade foram menos acuradas. Semelhantemente aos resultados obtidos quando as análises foram conduzidas com a população total, onde na maioria dos casos estudados, a variância genética aditiva foi subestimada (Quadro 9). As estimativas obtidas de 268,86, 88,02 e 91,88 ficaram abaixo dos valores simulados de 303,98, 191,77 e 140,46 para as subpopulações com alta, média e baixa variabilidade, respectivamente.

Pode-se observar que para as subpopulações pequenas em todos os tipos de heterogeneidades estudados as estimativas obtidas foram menos acuradas e em alguns níveis de variabilidade ocorreram problemas na convergência, ou não foram obtidas estimativas consistentes.

Quadro 8 - Valores dos componentes de variância genética aditiva e ambiental e de herdabilidade, das populações pequenas e grandes, com iguais níveis de variabilidade genética aditiva e ambiental, estimados pela metodologia REML, isoladamente.

Populações	$\hat{\sigma}_a^2$	$\hat{\sigma}_e^2$	\hat{h}^2	Nº de Indivíduos
HO-P-1	FC	FC	FC	100
HO-P-2	291,58	177,30	0,62	100
HO-P-3	258,50	236,07	0,52	100
HO-P-T	64,29	366,35	0,15	300
HO-G-1	269,17	297,57	0,47	1600
HO-G-2	301,89	266,29	0,53	1600
HO-G-3	371,66	195,97	0,65	1600
HO-G-T	370,73	209,83	0,64	4800

FC = Falhou a convergência.

Quadro 9 - Valores dos componentes de variância genética aditiva e ambiental e de herdabilidade, das subpopulações pequenas e grandes, sob três níveis de variabilidade genética aditiva, estimados pela metodologia REML, isoladamente.

Populações	$\hat{\sigma}_a^2$	$\hat{\sigma}_e^2$	\hat{h}^2	Nº de Indivíduos
HG-P-A	526,93	55,40	0,90	100
HG-P-M	208,43	237,98	0,47	100
HG-P-B	FC	FC	FC	100
HG-P-T	44,73	363,26	0,11	300
HG-G-A	268,86	297,68	0,47	1600
HG-G-M	88,02	353,36	0,20	1600
HG-G-B	91,88	378,04	0,24	1600
HG-G-T	201,79	274,57	0,42	4800

FC = Falhou a convergência.

Quadro 10 - Valores dos componentes de variância genética aditiva e ambiental e de herdabilidade, das subpopulações pequenas e grandes, sob três níveis de variabilidade ambiental, estimados pela metodologia REML, isoladamente.

Populações	$\hat{\sigma}_a^2$	$\hat{\sigma}_e^2$	\hat{h}^2	Nº de Indivíduos
HA-P-A	197,87	1648,88	0,11	100
HA-P-M	FC	FC	FC	100
HA-P-B	100,50	472,71	0,18	100
HA-P-T	323,43	841,40	0,28	300
HA-G-A	200,75	1404,83	0,13	1600
HA-G-M	98,09	1129,42	0,08	1600
HA-G-B	235,57	672,32	0,26	1600
HA-G-T	235,04	1021,89	0,19	4800

FC = Falhou a convergência.

Quadro 11 - Valores dos componentes de variância genética aditiva e ambiental e de herdabilidade das subpopulações pequenas e grandes, sob três níveis de variabilidade genética aditiva e ambiental, estimados pela metodologia REML, isoladamente.

Populações	$\hat{\sigma}_a^2$	$\hat{\sigma}_e^2$	\hat{h}^2	Nº de Indivíduos
HGA-P-A	71,46	1390,97	0,05	100
HGA-P-M	FC	FC	FC	100
HGA-P-B	65,96	473,96	0,12	100
HGA-P-T	17,94	1053,30	0,02	300
HGA-G-A	256,55	1518,64	0,14	1600
HGA-G-M	156,34	943,54	0,14	1600
HGA-G-B	67,99	531,65	0,11	1600
HGA-G-T	160,81	998,26	0,14	4800

FC = Falhou a convergência.

4.2.1.3. Porcentagens de Erros

Nos gráficos das Figuras de 5 a 8 são apresentadas as porcentagens de erros das estimativas obtidas pelo método REML, ao calcular os componentes de variância genética aditiva e ambiental, isoladamente para os níveis de variabilidade, oriundos das populações com diferentes tipos de heterogeneidade de variâncias e dois tamanhos de população. Os resultados são expressos em porcentagem de erro em relação aos valores reais dos componentes de variância.

De modo geral, é possível observar que para as populações de tamanho grande, independente do tipo de heterogeneidade estudado, as estimativas foram mais próximas dos valores reais, enquanto que, para as populações pequenas, em alguns casos, ocorreram problemas na convergência ou as estimativas produzidas foram inconsistentes. Isto pode ser explicado devido a uma redução no número de indivíduos ao realizar as análises isoladamente em cada nível de estratificação e calcular os componentes dentro de cada subclasse de variabilidade. Esta redução no número de indivíduos causou um problema na própria estrutura dos dados, afetando a sua normalidade, o que foi verificado em análises prévias com testes de normalidade.

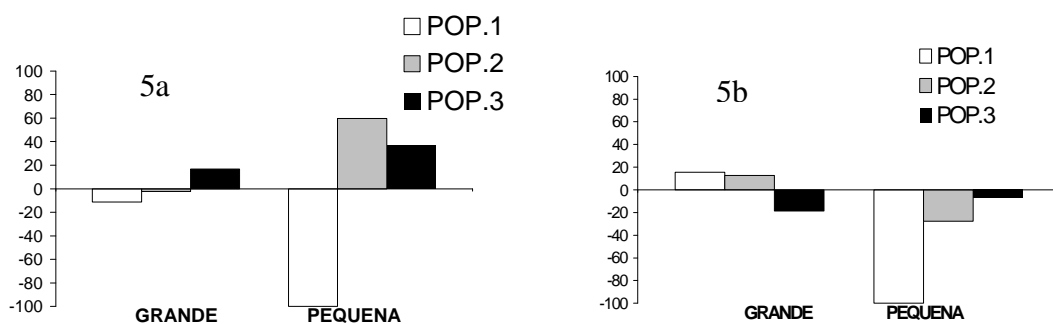


Figura 5. Percentagens de erro das estimativas dos componentes de variância genética aditiva (5a) e ambiental (5b), obtidos pela metodologia REML, em subpopulações com 100 indivíduos (PEQUENA) e subpopulações com 1600 indivíduos (GRANDE), com homogeneidade de variâncias.

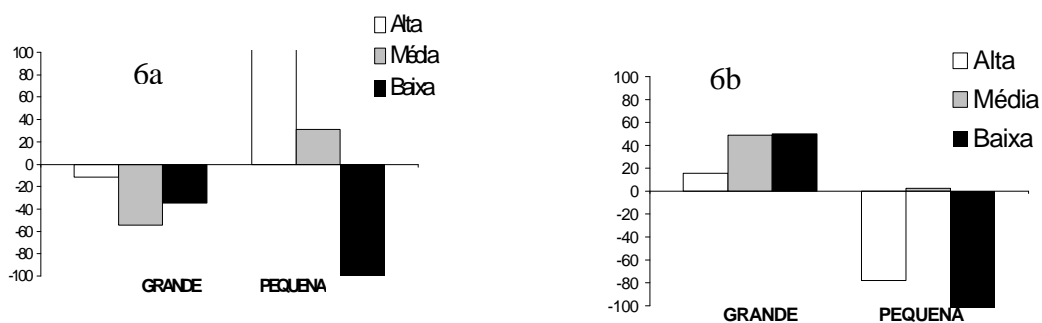


Figura 6. Percentagens de erro das estimativas dos componentes de variância genética aditiva (6a) e ambiental (6b), obtidos pela metodologia REML, em subpopulações com 100 indivíduos (PEQUENA) e subpopulações com 1600 indivíduos (GRANDE), oriundas da população com heterogeneidade de variância genética aditiva, sob três níveis de variabilidade.

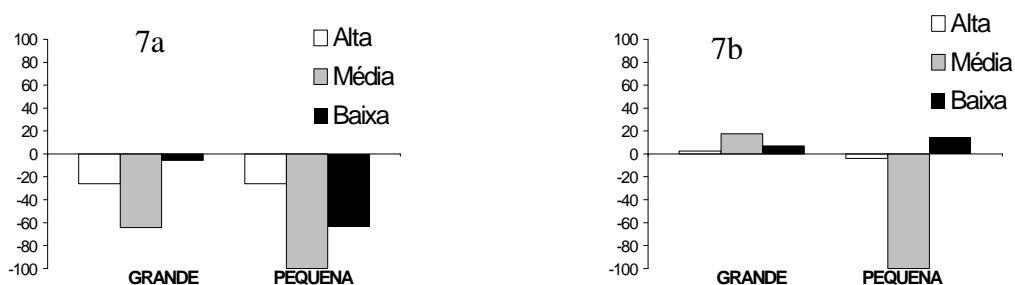


Figura 7. Percentagens de erro das estimativas dos componentes de variância genética aditiva (7a) e ambiental (7b), obtidos pela metodologia REML, em subpopulações com 100 indivíduos (PEQUENA) e subpopulações com 1600 indivíduos (GRANDE), oriundas da população com heterogeneidade de variância ambiental, sob três níveis de variabilidade.

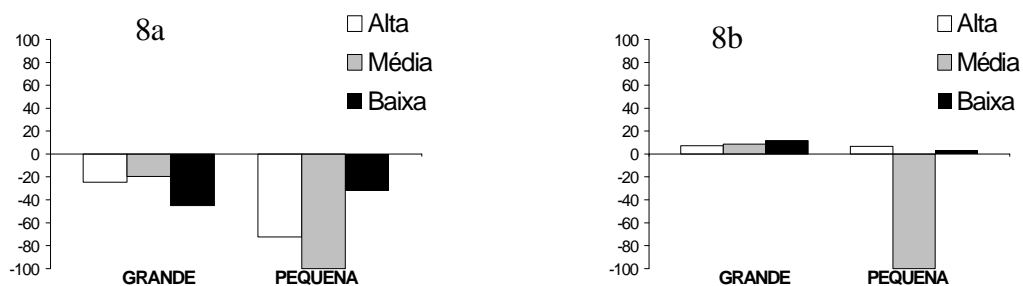


Figura 8. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (8a) e ambiental (8b), obtidos pela metodologia REML, em subpopulações com 100 indivíduos (PEQUENA) e subpopulações com 1600 indivíduos (GRANDE), oriundas da população com heterogeneidade de variância genética aditiva e ambiental, sob três níveis de variabilidade.

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 foram pequenas. Segundo WEIGEL e GIANOLA (1992), quando a heterogeneidade existe em rebanhos pequenos, a utilização de métodos que requerem justificação assintótica, tal qual o método REML, não produz estimativas confiáveis para componentes de variância dentro de rebanhos, concordando com as observações de WINKELMAN e SCHAEFFER (1988) que demonstraram ser a acurácia das estimativas dos componentes de variâncias influenciada quando as análises são realizadas dentro de subníveis de estratificação. Segundo esses autores, o pequeno número de animais por rebanho poderia conduzir a grandes erros de amostragens. Além disso, a convergência poderia não ocorrer para alguns rebanhos.

Como foi observado, as análises dentro dos subníveis formados também apresentaram problemas de convergência para as populações pequenas e em alguns casos as estimativas não foram consistentes.

4.2.2. Análises pela Metodologia BAYESIANA

4.2.2.1. Análises desconsiderando a Heterogeneidade

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

No Quadro 12 são apresentadas as estimativas dos componentes de variância genética aditiva e ambiental, obtidos pela metodologia Bayesiana, para as populações totais pequenas e grandes com homogeneidade de variâncias, heterogeneidade de variância genética aditiva, heterogeneidade de variância ambiental e heterogeneidade de variância genética aditiva e ambiental, considerando três níveis de informação. De forma geral, pode-se verificar uma diferença entre os valores obtidos para a moda e para a média nas populações de tamanho pequeno, quando *priors* pouco informativos foram utilizados. Estes resultados demonstram que a forma da distribuição *a posteriori* para estas situações afasta-se da normalidade, sendo que quando níveis mais informativos foram empregados, este fato não foi mais verificado.

Para a estrutura de população com homogeneidade de variâncias e heterogeneidade de variância ambiental, foi utilizada também uma distribuição normal contaminada para os resíduos (Quadro 13). Para esta situação foram considerados apenas *priors* não informativos. Os valores da média e da moda, neste caso, também foram diferentes entre si, semelhantes aos resultados obtidos quando foi utilizada normal Gaussiana.

Para melhor visualização 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 comparados com o nível de interesse: alta, média ou baixa variabilidade.

Quadro 12 - Valores dos componentes de variância genética aditiva e ambiental das populações totais pequenas e grandes, com homogeneidade de variâncias e presença de heterogeneidade de variância genética aditiva, heterogeneidade de variância ambiental e heterogeneidade de variância genética aditiva e ambiental, estimados pela metodologia Bayesiana.

Populações		\hat{S}_a^2			\hat{S}_e^2		
		NI	PI	I	NI	PI	I
HO-P-T	MÉDIA	72,91	113,21	176,52	361,83	329,24	281,30
	MODA	37,10	87,50	168,34	358,50	340,75	274,67
HO-G-T	MÉDIA	373,73	371,16	359,79	209,01	209,88	217,78
	MODA	372,92	373,00	357,15	204,22	212,32	221,88
HG-P-T	MÉDIA	58,66	98,82	158,57	361,33	328,11	281,51
	MODA	3,02	85,59	143,63	370,26	327,64	283,74
HG-G-T	MÉDIA	208,99	209,73	211,66	270,77	269,84	268,86
	MODA	206,92	215,28	213,83	267,87	266,69	266,14
HA-P-T	MÉDIA	379,55	274,76	262,93	819,92	886,34	928,48
	MODA	311,64	170,69	229,81	888,16	913,53	922,41
HA-G-T	MÉDIA	264,12	258,09	264,46	1003,28	1007,61	1001,72
	MODA	264,50	251,76	261,36	1014,10	1020,70	985,39
HGA-P-T	MÉDIA	52,59	108,40	155,39	1034,35	980,42	954,65
	MODA	0,35	66,54	142,21	984,49	993,41	956,32
HGA-G-T	MÉDIA	182,36	180,02	194,80	983,29	985,90	973,79
	MODA	169,42	184,33	196,33	978,98	985,59	982,80

NI= Não informativo; PI = Pouco informativo; I = Informativo.

Quadro 13 - Valores dos componentes de variância genética aditiva e ambiental das populações totais pequenas e grandes, com homogeneidade de variâncias e presença de heterogeneidade de variância ambiental, estimados pela metodologia Bayesiana, considerando distribuição normal contaminada para os resíduos.

Populações		\hat{S}_a^2	\hat{S}_e^2
		NI	NI
HO-P-T	MÉDIA	87,09	354,05
	MODA	57,56	369,12
HO-G-T	MÉDIA	373,19	209,59
	MODA	372,71	206,06
HA-P-T	MÉDIA	422,10	796,84
	MODA	246,76	851,08
HA-G-T	MÉDIA	245,09	1014,99
	MODA	224,63	1025,53

NI= Não informativo

4.2.2.2. Porcentagens de Erros

Nos Gráficos das Figuras 9 a 14, são apresentadas as porcentagens de erros obtidas pelo método Bayesiano, no cálculo dos componentes de variância genética aditiva e ambiental, quando a heterogeneidade foi desconsiderada. De modo geral, as populações pequenas tiveram maiores problemas na estimação. Para as populações com homogeneidade de variâncias (Figura 9), pode ser observado que a introdução de maiores níveis de informação resultou em menores porcentagens de erros, tanto para a variância genética aditiva quanto para a variância ambiental. Entretanto, para as estruturas de heterogeneidade analisadas o aumento no nível de informação *a priori* só melhorou a estimação do componente de variância não heterogêneo. 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 obtido foi quando a população era de tamanho grande e apresentava homogeneidade de variâncias (Figuras 9c e 9d). Para esta situação, o aumento no nível de informação não foi tão expressivo quanto para a população pequena (Figuras 9a e 9b), demonstrando que a importância da informação *a priori* aumenta em populações de tamanho pequeno. Para a metodologia Bayesiana foi utilizada também uma distribuição normal contaminada com *priors* não informativos para os resíduos, esta forma de análise foi empregada para a população com homogeneidade de variâncias e para a população com heterogeneidade de variância ambiental. Nas Figuras 10 e 13 são apresentadas as porcentagens de erros, na situação com homogeneidade de variâncias e heterogeneidade de variância ambiental para as três opções de análise: metodologia REML; metodologia Bayesiana com distribuição normal Gaussiana para os resíduos, sob flat *priors* (GAUSS-NI); e metodologia Bayesiana com distribuição normal contaminada para os resíduos sob flat *priors* (NC-NI).

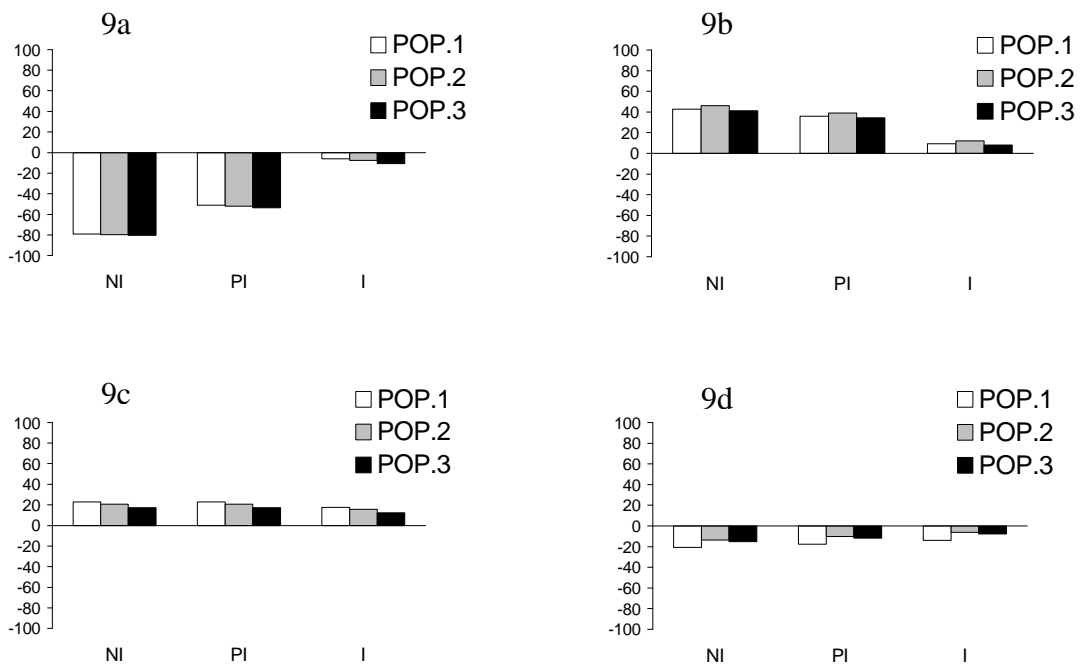


Figura 9. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (9a, 9c) e ambiental (9b, 9d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em populações com 300 indivíduos (9a, 9b) e populações com 4800 indivíduos (9c, 9d), com homogeneidade de variâncias.

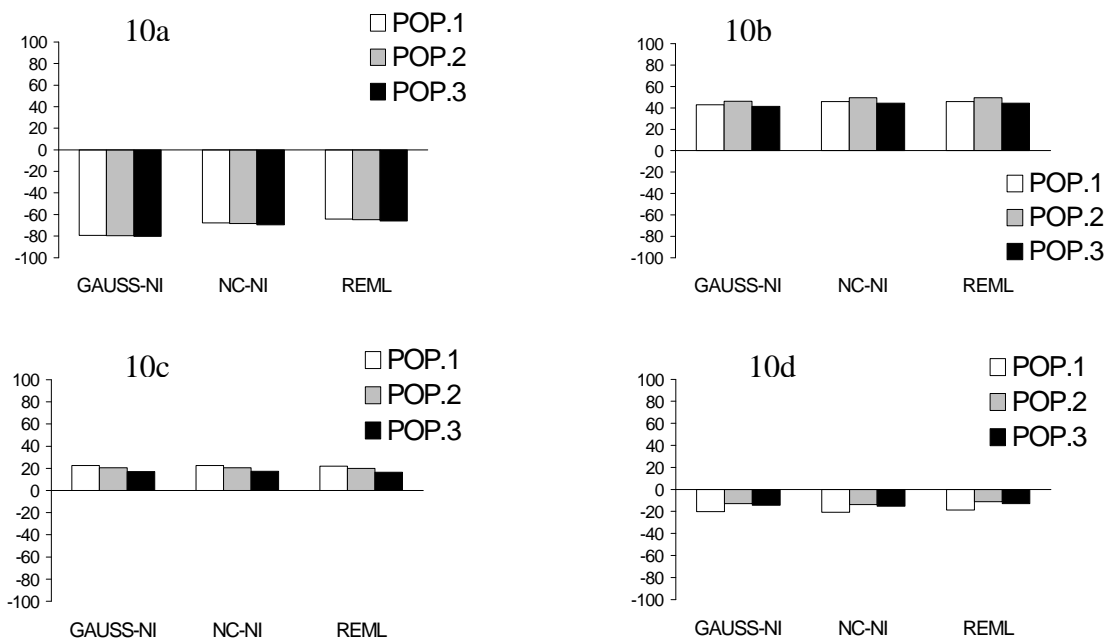


Figura 10. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (10a, 10c) e ambiental (10b, 10d), obtidos pela metodologia Bayesiana, sob distribuição normal Gaussiana e *prior* não informativo (GAUSS - NI); e distribuição normal Contaminada com *prior* não informativo (NC-NI); e pela metodologia REML, em populações com 300 indivíduos (10a, 10b) e populações com 4800 indivíduos (10c, 10d), com homogeneidade de variâncias.

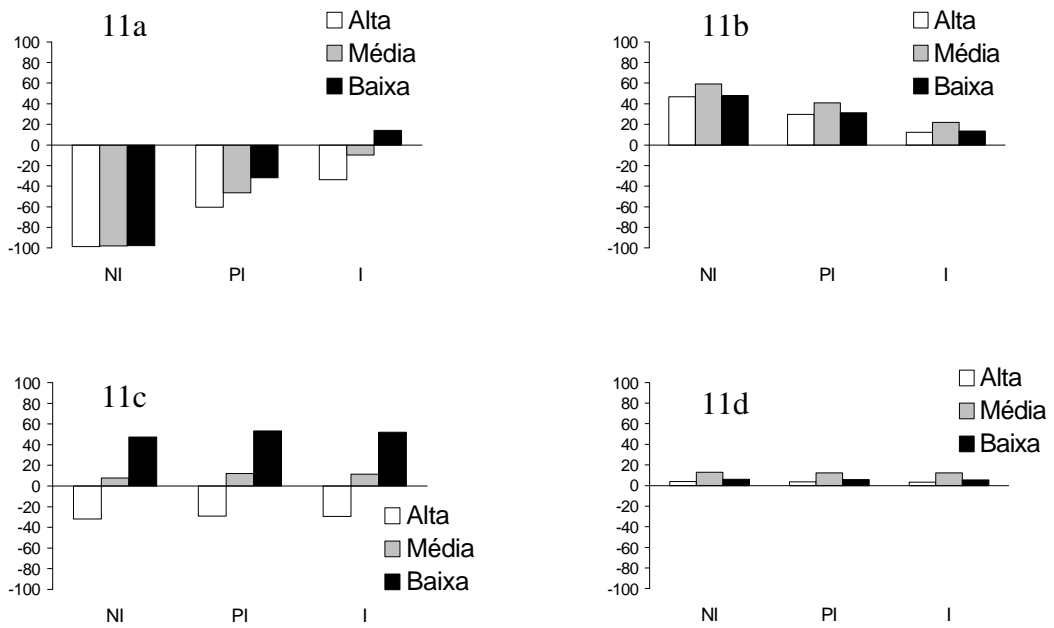


Figura 11. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (11a, 11c) e ambiental (11b, 11d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em populações com 300 indivíduos (11a, 11b) e populações com 4800 indivíduos (11c, 11d), com heterogeneidade de variância genética aditiva com três níveis de variabilidade.

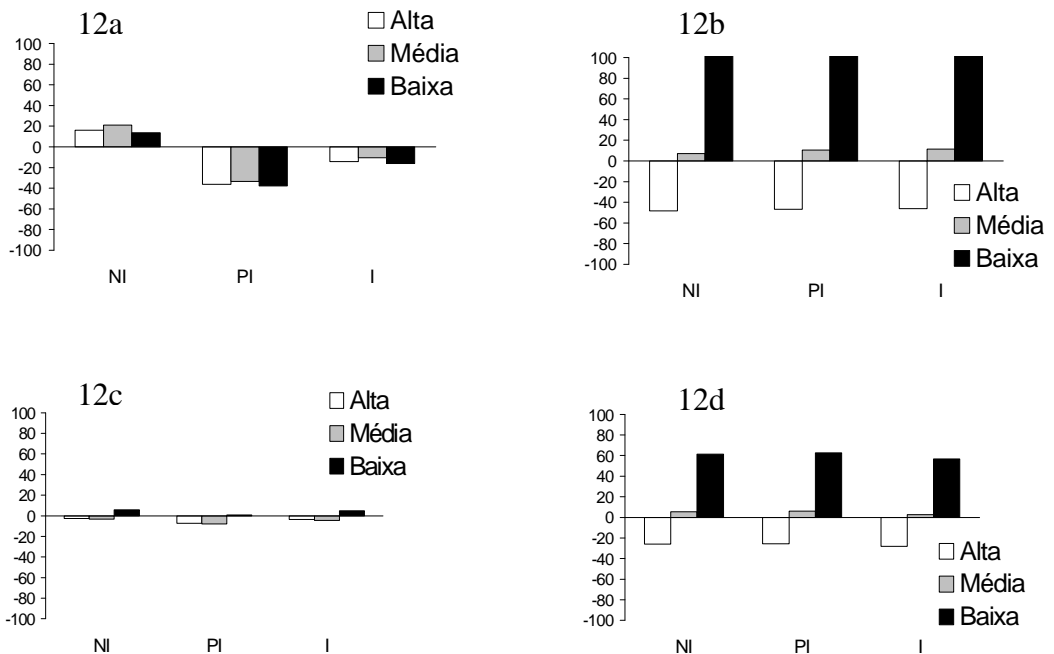


Figura 12. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (12a, 12c) e ambiental (12b, 12d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em populações com 300 indivíduos (12a, 12b) e populações com 4800 indivíduos (12c, 12d), com heterogeneidade de variância ambiental com três níveis de variabilidade.

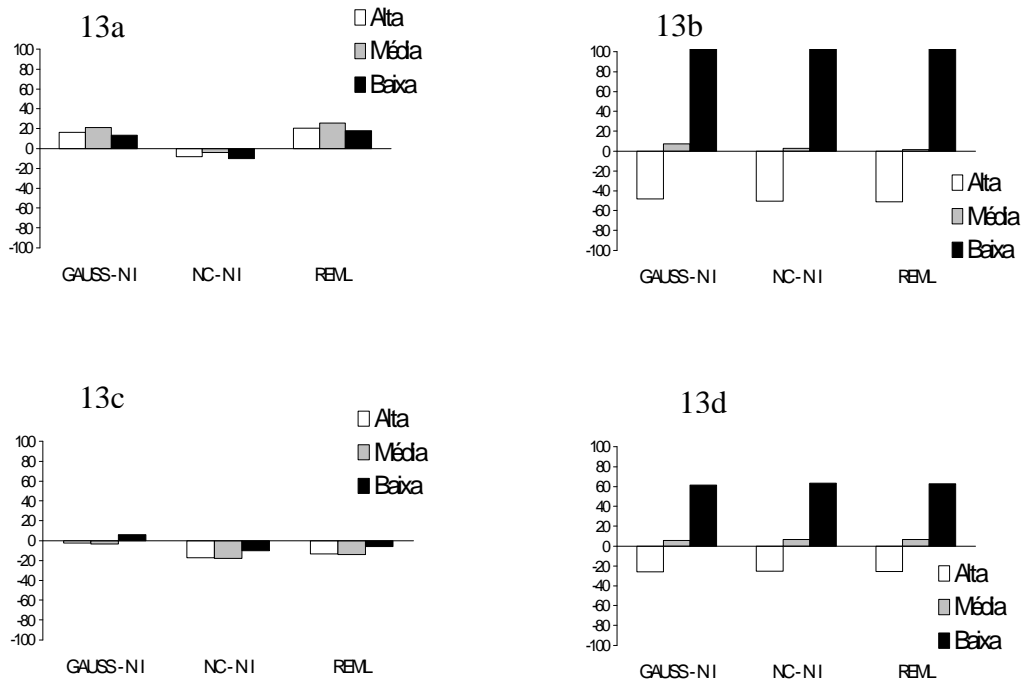


Figura 13. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (13a, 13c) e ambiental (13b, 13d), obtidos pela metodologia Bayesiana, sob distribuição normal Gaussiana e *prior* não informativo (GAUSS - NI); e distribuição normal Contaminada com *prior* não informativo (NC-NI) e pela metodologia REML, em populações com 300 indivíduos (13a, 13b) e populações com 4800 indivíduos (13c, 13d), com heterogeneidade de variância ambiental com três níveis de variabilidade.

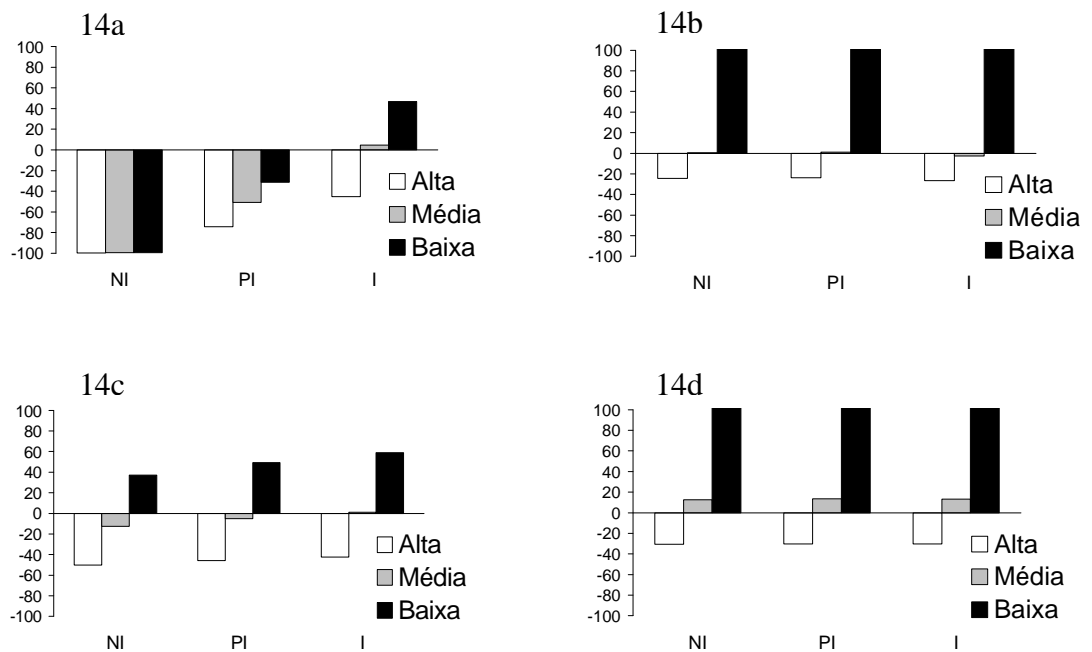


Figura 14. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (14a, 14c) e ambiental (14b, 14d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em populações com 300 indivíduos (14a, 14b) e populações com 4800 indivíduos (14c, 14d), com heterogeneidade de variância genética aditiva e ambiental com três níveis de variabilidade.

Na situação onde a presença da heterogeneidade genética aditiva foi desconsiderada (Figura 11), até mesmo a população de tamanho grande apresentou problemas na estimação da variância genética aditiva, sendo que o nível de informação *a priori* não foi capaz de melhorar a estimação. Entretanto, a variância ambiental não foi influenciada pela heterogeneidade genética aditiva. Quando a heterogeneidade foi no componente de variância ambiental (Figura 12), a variância genética aditiva foi bem estimada e a variância ambiental apresentou maiores problemas. Contudo, as subpopulações com variabilidade média não apresentaram problemas pelo fato de seus valores de variância ambiental estar próximos do valor médio dos três níveis. A utilização da distribuição normal contaminada, para esta situação não foi eficiente em diminuir as porcentagens de erros (Figura 13). Portanto, quando a heterogeneidade é desconsiderada mesmo utilizando-se um método robusto de análise, alguns níveis de variabilidade podem ter seus valores subestimados ou superestimados.

Na estrutura de população que apresenta heterogeneidade nos dois componentes de variância (Figura 14), tanto a variância genética aditiva quanto a variância ambiental, tiveram seus valores subestimados para o nível de alta variabilidade e superestimados para o nível de baixa variabilidade, sendo que a informação *a priori* não produziu nenhum efeito satisfatório, com exceção da população pequena (Figura 14a), onde a convergência não foi obtida quando se utilizou *prior* não informativo.

Na tentativa de considerar a presença de heterogeneidade de variâncias, com a metodologia Bayesiana, foi utilizado um modelo multicaracterístico. Contudo, apesar da convergência ter sido alcançada, os resultados obtidos ficaram bem distantes dos valores simulados. Procederam-se desta forma, as análises dentro dos subníveis.

4.2.2.3. Análises dentro dos Subníveis

Como o modelo multicaracterístico não se mostrou adequado, optou-se pelas análises de característica única, considerando cada subpopulação isoladamente. As estimativas dos componentes de variância genética aditiva e ambiental, obtidas pela metodologia Bayesiana com diferentes níveis de informações *a priori*, considerando cada ambiente isoladamente, são apresentados nos Quadros 14 a 17. Para as subpopulações oriundas das populações com heterogeneidade ambiental foi utilizada também a metodologia Bayesiana, considerando distribuição normal contaminada para os resíduos. Os valores das médias e das modas, para esta situação, são apresentados nos Quadros 18 e 19.

Pode-se observar que para as populações pequenas em todos os tipos de heterogeneidade estudados e em alguns níveis de variabilidade, 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. Isto pode ser explicado pelo fato da normalidade ter sido considerada *a priori* para os valores genéticos dos indivíduos. Talvez se tivesse sido utilizada uma distribuição robusta para os valores genéticos, que fosse menos sensível ao desvio da normalidade, as estimativas poderiam apresentar-se mais consistentes. Entretanto, quando maiores níveis de informação foram utilizados, as estimativas se aproximaram dos valores reais. Estes resultados estão de acordo com os obtidos por VAN TASSEL et al. (1995) que concluíram que a utilização da informação *a priori* pode contribuir para a estimação dos componentes de variância.

Quadro 14 - Valores da média e da moda da distribuição *a posteriori* dos componentes de variância genética aditiva e ambiental, das populações pequenas e grandes, com iguais níveis de variabilidade genética aditiva e ambiental, estimados sob três níveis de informação *a priori*, pela metodologia Bayesiana.

Populações		\hat{S}_a^2			\hat{S}_e^2		
		NI	PI	I	NI	PI	I
HO-P-1	MÉDIA	127,92	127,92	174,46	301,68	301,68	260,75
	MODA	84,37	84,37	163,57	290,24	290,24	254,38
HO-P-2	MÉDIA	13,29	189,66	188,52	393,63	252,48	246,68
	MODA	0,33	165,68	181,71	378,06	275,50	239,11
HO-P-3	MÉDIA	449,39	207,53	198,61	93,30	269,35	261,02
	MODA	554,35	119,76	181,99	0,41	266,95	248,11
HO-P-T	MÉDIA	72,91	113,21	176,52	361,83	329,24	281,30
	MODA	37,10	87,50	168,34	358,50	340,75	274,67
HO-G-1	MÉDIA	284,83	279,83	294,23	288,22	289,93	277,32
	MODA	266,41	277,56	286,21	302,99	296,59	286,80
HO-G-2	MÉDIA	317,21	314,74	320,69	256,63	257,69	251,09
	MODA	318,69	308,65	315,14	275,97	254,70	248,86
HO-G-3	MÉDIA	382,73	364,02	335,88	190,84	203,06	225,97
	MODA	370,71	379,75	334,04	192,23	203,72	223,91
HO-G-T	MÉDIA	373,73	371,16	359,79	209,01	209,88	217,78
	MODA	372,92	373,00	357,15	204,22	212,32	221,88

NI= Não Informativo; PI = Pouco Informativo; I = Informativo.

Quadro 15 - Valores da média e da moda da distribuição *a posteriori* dos componentes de variância genética aditiva e ambiental, das populações pequenas e grandes, com três níveis de variabilidade genética aditiva, estimados sob três níveis de informação *a priori*, pela metodologia Bayesiana.

Populações		\hat{S}_a^2			\hat{S}_e^2		
		NI	PI	I	NI	PI	I
HG-P-A	MÉDIA	542,90	274,74	232,69	32,58	228,49	257,10
	MODA	546,82	196,13	211,99	0,27	245,07	247,49
HG-P-M	MÉDIA	270,52	181,44	168,85	207,63	252,14	246,27
	MODA	0,80	78,58	156,01	0,47	303,61	240,31
HG-P-B	MÉDIA	39,96	86,15	119,48	288,64	254,34	243,99
	MODA	0,57	49,37	110,03	298,61	245,91	233,15
HG-P-T	MÉDIA	58,66	98,82	158,57	361,33	328,11	281,51
	MODA	3,02	85,59	143,63	370,26	327,64	283,74
HG-G-A	MÉDIA	281,13	281,59	292,74	291,28	289,91	278,61
	MODA	267,78	279,33	285,81	272,68	289,55	289,49
HG-G-M	MÉDIA	104,46	112,17	299,42	340,02	337,58	299,42
	MODA	95,14	107,22	299,57	329,77	354,35	299,57
HG-G-B	MÉDIA	110,36	108,42	125,69	277,49	274,21	260,61
	MODA	109,68	98,38	120,94	276,20	275,49	258,91
HG-G-T	MÉDIA	208,99	209,73	211,66	270,77	269,84	268,86
	MODA	206,92	215,28	213,83	267,87	266,69	266,14

NI= Não Informativo; PI = Pouco Informativo; I = Informativo.

Quadro 16 - Valores da média e da moda da distribuição *a posteriori* dos componentes de variância genética aditiva e ambiental, das populações pequenas e grandes, com três níveis de variabilidade ambiental, estimados sob três níveis de informação *a priori*, pela metodologia Bayesiana.

Populações		\hat{S}_a^2			\hat{S}_e^2		
		NI	PI	I	NI	PI	I
HA-P-A	MÉDIA	641,33	257,78	264,69	1107,10	1615,10	1673,91
	MODA	41,91	131,76	250,57	1711,01	1572,05	1629,47
HA-P-M	MÉDIA	98,89	181,36	242,74	893,83	784,34	790,78
	MODA	1,34	115,51	213,56	859,50	804,01	756,21
HA-P-B	MÉDIA	166,47	193,12	251,28	1748,02	398,22	393,69
	MODA	2,57	170,08	235,55	1662,18	385,70	387,33
HA-P-T	MÉDIA	379,55	274,76	262,93	819,92	886,34	928,48
	MODA	311,64	170,69	229,81	888,16	913,53	922,41
HA-G-A	MÉDIA	156,83	250,78	272,59	1365,62	1371,06	1349,36
	MODA	169,79	232,85	255,86	1348,70	1359,22	1364,87
HA-G-M	MÉDIA	134,19	166,61	238,17	1045,51	1075,48	1011,00
	MODA	133,26	127,15	243,06	1038,39	1087,87	1004,79
HA-G-B	MÉDIA	211,03	252,69	258,29	659,49	658,95	654,12
	MODA	192,12	259,84	253,87	662,20	674,63	662,41
HA-G-T	MÉDIA	264,12	258,09	264,46	1003,28	1007,61	1001,72
	MODA	264,50	251,76	261,36	1014,10	1020,70	985,39

NI= Não Informativo; PI = Pouco Informativo; I = Informativo.

Quadro 17 - Valores da média e da moda da distribuição *a posteriori* dos componentes de variância genética aditiva e ambiental, das populações pequenas e grandes, com três níveis de variabilidade genética e ambiental, estimados sob três níveis de informação *a priori*, pela metodologia Bayesiana.

Populações		\hat{S}_a^2			\hat{S}_e^2		
		NI	PI	I	NI	PI	I
HGA-P-A	MÉDIA	164,93	217,92	242,93	1310,46	1264,96	1284,83
	MODA	1,30	113,70	190,60	1313,95	1161,93	1251,26
HGA-P-M	MÉDIA	48,56	105,38	133,98	1157,60	1087,24	1013,15
	MODA	1,16	62,46	121,64	1151,70	995,94	1007,94
HGA-P-B	MÉDIA	67,21	88,18	95,69	497,62	458,63	455,73
	MODA	0,77	52,82	85,07	489,10	445,00	451,13
HGA-P-T	MÉDIA	52,59	108,40	155,39	1034,35	980,42	954,65
	MODA	0,35	66,54	142,21	984,49	993,41	956,32
HGA-G-A	MÉDIA	312,93	300,86	331,87	1467,29	1474,47	1455,62
	MODA	282,03	304,65	313,26	1481,36	1502,36	1448,90
HGA-G-M	MÉDIA	201,96	183,79	197,00	898,37	914,52	905,35
	MODA	183,88	167,01	175,51	940,51	908,87	901,38
HGA-G-B	MÉDIA	96,84	94,10	115,62	508,97	511,40	490,31
	MODA	76,49	87,01	113,53	510,59	521,25	501,83
HGA-G-T	MÉDIA	182,36	180,02	194,80	983,29	985,90	973,79
	MODA	169,42	184,33	196,33	978,98	985,59	982,80

NI= Não informativo; PI = Pouco informativo; I = Informativo.

Quadro 18 - Valores da média e da moda da distribuição *a posteriori* dos componentes de variância genética aditiva e ambiental, das populações pequenas e grandes com iguais níveis de variabilidade genética aditiva e ambiental, estimados pela metodologia Bayesiana, utilizando a distribuição normal contaminada para os resíduos.

Populações		\hat{S}_a^2	\hat{S}_e^2
		NI	NI
HO-P-1	MÉDIA	125,35	326,74
	MODA	9,48	326,66
HO-P-2	MÉDIA	314,41	195,93
	MODA	374,75	221,68
HO-P-3	MÉDIA	329,61	224,23
	MODA	297,19	229,23
HO-P-T	MÉDIA	87,09	354,05
	MODA	57,56	369,12
HO-G-1	MÉDIA	275,39	291,73
	MODA	257,86	295,17
HO-G-2	MÉDIA	301,73	268,74
	MODA	280,54	271,00
HO-G-3	MÉDIA	367,61	198,55
	MODA	396,15	209,65
HO-G-T	MÉDIA	373,19	209,59
	MODA	372,71	206,06

NI = Não Informativo

Quadro 19 - Valores das médias e das modas das distribuições *a posteriori* dos componentes de variância genética aditiva e ambiental, das populações pequenas e grandes, com três níveis de variabilidade ambiental, estimados pela metodologia Bayesiana, utilizando a distribuição normal contaminada para os resíduos.

Populações		\hat{S}_a^2	\hat{S}_e^2
		NI	NI
HA-P-A	MÉDIA	817,72	1283,20
	MODA	25,43	1598,57
HA-P-M	MÉDIA	322,01	742,44
	MODA	1,36	814,64
HA-P-B	MÉDIA	212,68	411,57
	MODA	79,95	483,73
HA-P-T	MÉDIA	422,10	796,84
	MODA	246,76	851,08
HA-G-A	MÉDIA	247,61	1360,65
	MODA	232,40	1426,60
HA-G-M	MÉDIA	131,14	1110,68
	MODA	127,64	1109,04
HA-G-B	MÉDIA	251,50	666,09
	MODA	241,21	685,37
HA-G-T	MÉDIA	245,09	1014,99
	MODA	224,63	1025,53

NI = Não Informativo

4.2.2.3. Porcentagens de Erros

Nos gráficos das Figuras de 15 a 18 são apresentadas as porcentagens de erros das estimativas obtidas pelo método Bayesiano, ao calcular os componentes de variância genética aditiva e ambiental, quando as subpopulações foram analisadas isoladamente, utilizando-se três níveis de informação *a priori*.

Nos gráficos da Figura 15 são apresentadas as porcentagens de erros das estimativas obtidas das subpopulações controle, oriundas da população controle com homogeneidade de variâncias. Observa-se que a subpopulação de tamanho grande apresentou estimativas próximas dos valores reais dos componentes de variâncias, sendo pouco influenciada pelo nível de informação *a priori*. Entretanto, a subpopulação pequena foi fortemente influenciada pela informação *a priori*. Este resultado também pode ser observado nas demais subpopulações, oriundas das populações com presença de heterogeneidade (Figuras 16, 17 e 18), concordando com os resultados de diversos autores, entre eles WANG et al. (1994) e BLASCO et al. (1998), em que, 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.

No gráfico da Figura 19 são apresentadas as porcentagens de erros das estimativas obtidas para as subpopulações oriundas da população com heterogeneidade de variância ambiental, quando foi utilizada distribuição normal contaminada para os resíduos. Pode-se observar que a metodologia REML e a metodologia Bayesiana com distribuição normal e com distribuição normal contaminada para os resíduos apresentaram porcentagens de erros semelhantes entre si, com exceção do componente de variância ambiental para a população pequena e variabilidade média, onde a utilização da normal contaminada conduziu a menores porcentagens de erros (Figura 19b).

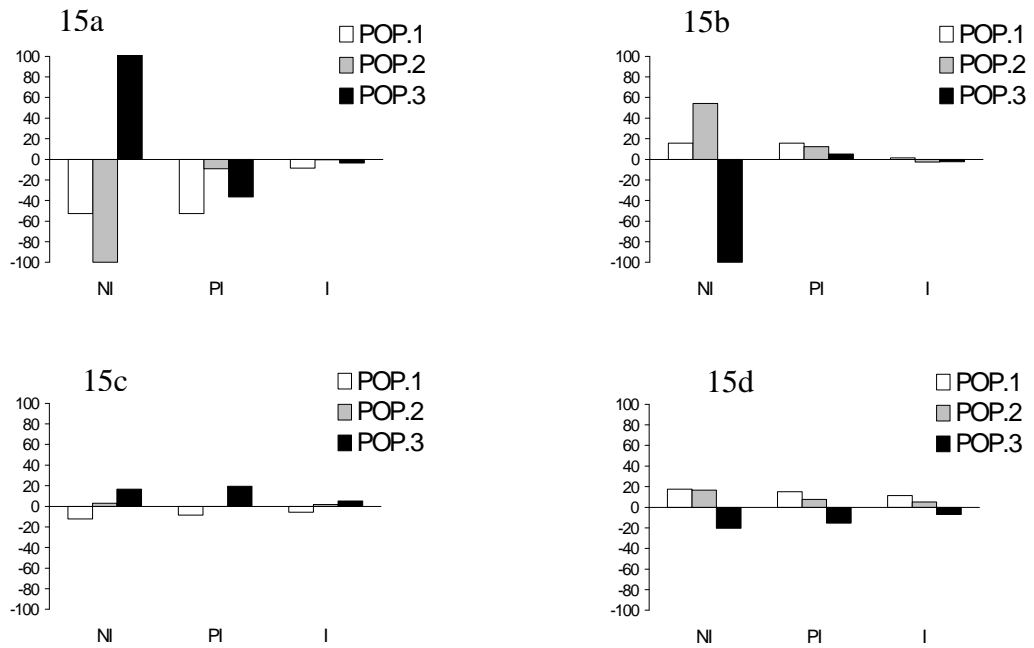


Figura 15. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (15a, 15c) e ambiental (15b, 15d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em subpopulações com 100 indivíduos (15a, 15b) e subpopulações com 1600 indivíduos (15c, 15d), provenientes da população com homogeneidade de variâncias.

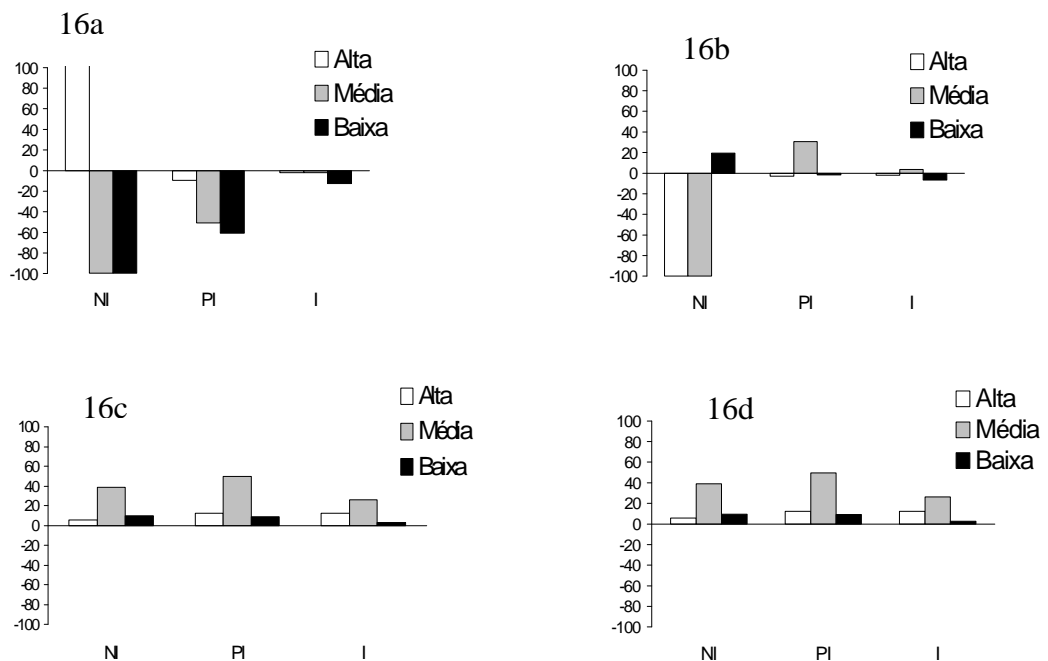


Figura 16. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (16a, 16c) e ambiental (16b, 16d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em subpopulações com 100 indivíduos (16a, 16b) e subpopulações com 1600 indivíduos (16c, 16d), provenientes da população com heterogeneidade genética.

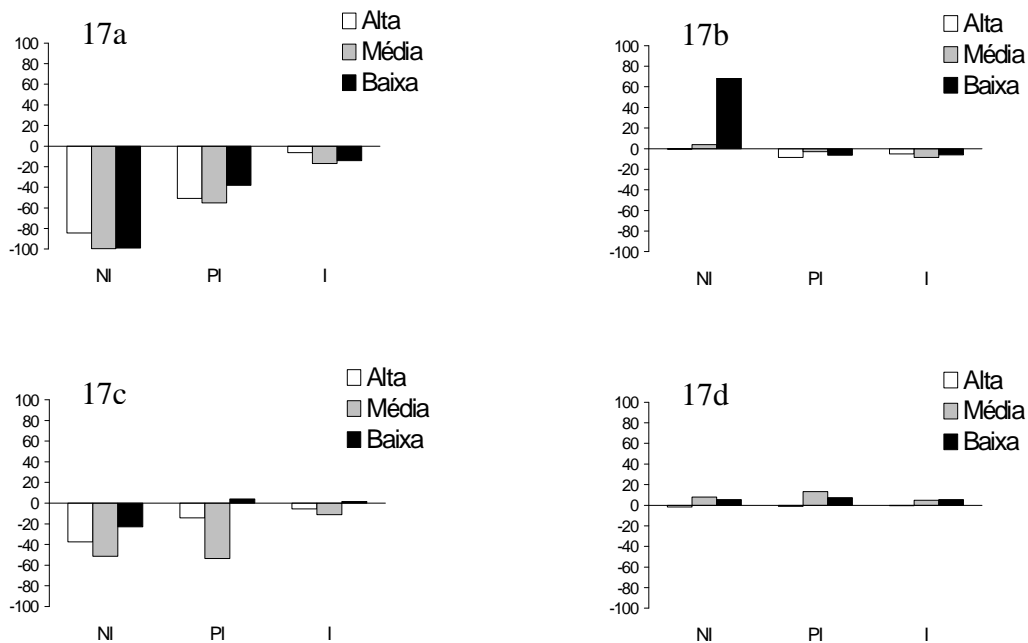


Figura 17. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (17a, 17c) e ambiental (17b, 17d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em subpopulações com 100 indivíduos (17a, 17b) e subpopulações com 1600 indivíduos (17c, 17d), provenientes da população com heterogeneidade ambiental.

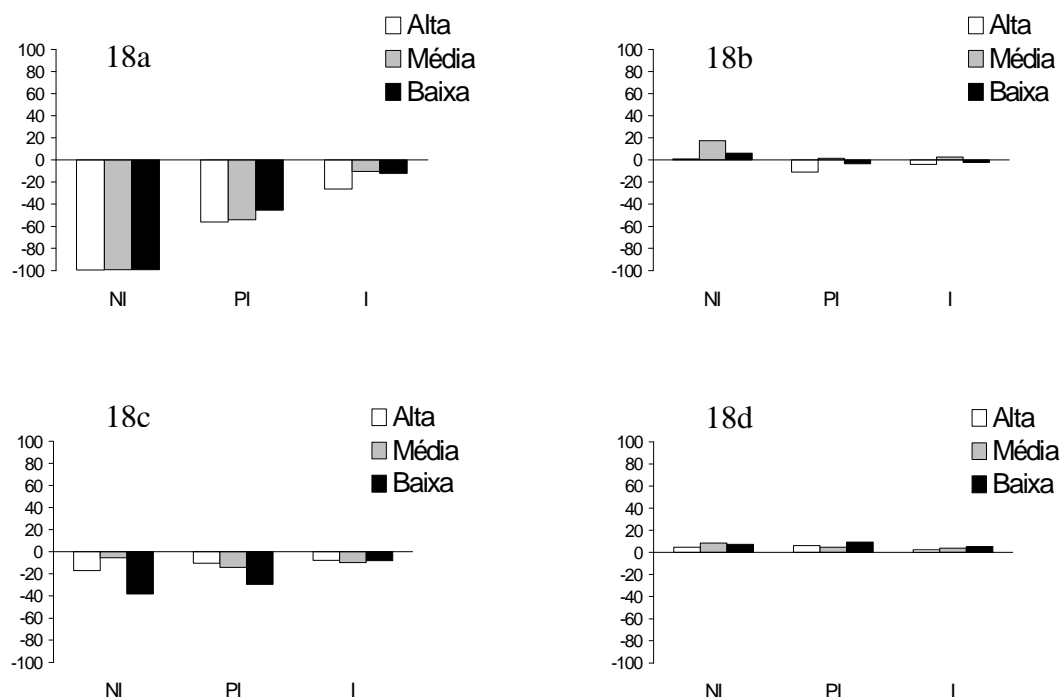


Figura 18. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (18a, 18c) e ambiental (18b, 18d), obtidos pela metodologia Bayesiana, sob três níveis de informação *a priori*: não informativo (NI); pouco informativo (PI) e informativo (I), em subpopulações com 100 indivíduos (18a, 18b) e subpopulações com 1600 indivíduos (18c, 18d), provenientes da população com heterogeneidade genética aditiva e ambiental.

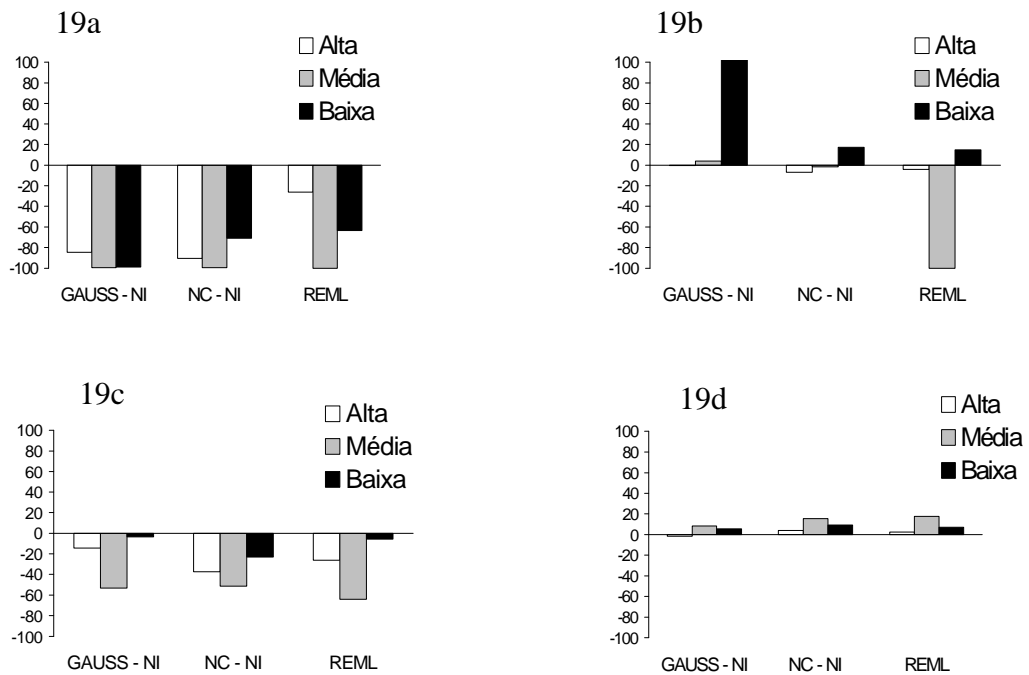


Figura 19. Porcentagens de erro das estimativas dos componentes de variância genética aditiva (19a, 19c) e ambiental (19b, 19d), obtidos pela metodologia REML e metodologia Bayesiana, sob distribuição normal Gaussiana e *prior* não informativo (GAUSS - NI); e distribuição normal Contaminada com *prior* não informativo (NC-NI), em subpopulações com 100 indivíduos (19a, 19b) e subpopulações com 1600 indivíduos (19c,19d), provenientes da população com heterogeneidade ambiental.

Conforme esperado, as estimativas produzidas pelo método REML e pelo método Bayesiano, com *priors* não informativos, foram semelhantes entre si. A vantagem do método Bayesiano, neste caso, foi a utilização de informações *a priori* nas análises realizadas dentro dos subníveis, principalmente em subpopulações de tamanho pequeno, indicando que a informação *a priori* é relativamente mais importante quando os dados são escassos, de acordo com GIANOLA e FERNANDO (1986).

Segundo WEIGEL e GIANOLA (1992), o conhecimento deste fato pode ser utilizado para a estimação de componentes de variância, combinando-se informações obtidas entre rebanhos e dentro de rebanhos, resultando desta forma, em estimativas menos influenciadas pela heterogeneidade de variâncias.

4.3. PREDIÇÃO DOS VALORES GENÉTICOS

4.3.1. Correlação de Ordem – SPEARMAN

Como a simulação foi realizada em nível genético, os valores genéticos reais dos indivíduos eram conhecidos, o que permitiu uma comparação da classificação dos indivíduos com base nos valores genéticos reais e nos valores genéticos preditos, na presença dos diferentes tipos de heterogeneidade de variâncias. Portanto, foi empregada a correlação de ordem entre os indivíduos, utilizando os valores genéticos preditos pelo método freqüentista REML – BLUP, denominado BLUP empírico ou EBLUP (via Equações de Modelos Mistos de Henderson) e pelo método Bayesiano.

No Quadro 20 são apresentadas as correlações de ordem quando a presença dos diferentes tipos de heterogeneidade foi desconsiderada. Em todas as situações os melhores resultados foram obtidos com as populações de tamanho grande. A população grande com homogeneidade de variâncias apresentou a maior correlação de ordem 0,78, sendo que os métodos foram similares entre si e o aumento no nível de informação não melhorou a classificação dos indivíduos. A utilização da distribuição normal contaminada também não ocasionou mudança na classificação dos indivíduos para a população grande. Entretanto, para a população pequena, com homogeneidade de variâncias, o aumento no nível de informação aumentou a correlação de 0,61, com flat *priors*, para 0,66 no nível informativo, sendo que o valor obtido com a distribuição normal contaminada de 0,63 foi superior ao obtido pelo método EBLUP de 0,59.

Pode-se verificar que para as demais estruturas de população, menores correlações de ordem foram obtidas. Para a população pequena com heterogeneidade genética, o aumento no nível de informação também ocasionou melhor classificação dos indivíduos, sendo que o valor obtido pelo método Bayesiano, com flat *priors* de 0,60, foi superior ao obtido pelo método EBLUP de 0,57, contudo este resultado não é suficiente para maiores conclusões. As menores correlações de ordem foram encontradas quando a heterogeneidade ambiental foi desconsiderada, sendo de 0,40 para a população pequena e de 0,54 para a população grande. Para esta situação o

aumento do nível de informação não ocasionou melhoras na classificação dos indivíduos, tanto a população grande quanto para a população pequena.

Quadro 20 - Correlação de ordem entre os valores genéticos reais e preditos, obtidos pela metodologia EBLUP e Bayesiana, com distribuição normal, sob três níveis de informação *a priori* e distribuição normal contaminada (NC), com flat *priors*, desconsiderando a presença de heterogeneidade de variâncias.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HO-P-T	0,59	0,61	0,64	0,66	0,63
HO-G-T	0,78	0,78	0,78	0,78	0,78
HG-P-T	0,57	0,60	0,64	0,66	-
HG-G-T	0,71	0,72	0,72	0,72	-
HA-P-T	0,40	0,40	0,40	0,40	0,40
HA-G-T	0,54	0,54	0,54	0,54	0,54
HGA-P-T	0,42	0,43	0,43	0,44	-
HGA-G-T	0,54	0,54	0,54	0,54	-

NI= Não informativo; PI = Pouco informativo; I = Informativo, NC = Normal contaminada.

Os valores obtidos para as populações com heterogeneidade ambiental e heterogeneidade genética e ambiental foram similares e bem inferiores aos obtidos para a estrutura com homogeneidade de variâncias e heterogeneidade genética, tanto em populações grandes quanto em pequenas, para ambos os métodos. A utilização da normal contaminada também conduziu a valores semelhantes.

Ressalta-se desta forma, que a presença da heterogeneidade de variância reduz a eficiência da classificação dos indivíduos, principalmente se a heterogeneidade ocorrer no componente ambiental, indicando que maiores cuidados devem ser tomados na pradonização de práticas de manejo, nutrição dos animais e aspectos sanitários dos rebanhos. Portanto, desconsiderar a presença de heterogeneidade, mesmo para populações grandes, pode resultar em prejuízo para a classificação dos animais, causando diminuição no progresso genético. Diferentemente das observações de WELLER (1985) e BOLDMAN e FREEMAN (1990) que, trabalhando com dados reais, compararam análises que consideraram e que desconsideraram a heterogeneidade, encontrando correlações de ordem entre as duas formas de avaliação próximas de 0,99.

Com relação às metodologias empregadas, não foram encontradas diferenças consideráveis, e ambas apresentaram perdas semelhantes ao se desconsiderar a presença de heterogeneidade. Para verificar o efeito de se desconsiderar a heterogeneidade dentro dos níveis de variabilidade, os resultados obtidos a partir das populações totais foram comparados com os valores genéticos reais dos subníveis formados. As correlações de ordem para esta situação são apresentadas nos Quadros 21 a 23.

Quadro 21 – Correlação de ordem entre os valores genéticos reais e preditos, obtidos pela metodologia EBLUP e Bayesiana, sob três níveis de informação, em populações com três níveis de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância genética aditiva.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HG-P-A	0,63	0,62	0,64	0,64
HG-P-M	0,56	0,56	0,55	0,55
HG-P-B	FC	0,40	0,38	0,38
HG-G-A	0,68	0,69	0,69	0,69
HG-G-M	0,62	0,63	0,63	0,64
HG-G-B	0,51	0,51	0,51	0,52

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

Quadro 22 – Correlação de ordem entre os valores genéticos reais e preditos, obtidos pela metodologia EBLUP e Bayesiana com distribuição normal, sob três níveis de informação e distribuição normal contaminada (NC), com flat priors, em populações com três níveis de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância ambiental.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HA-P-A	0,40	0,37	0,38	0,38	0,34
HA-P-M	FC	0,36	0,36	0,37	0,34
HA-P-B	0,32	0,31	0,32	0,32	0,54
HA-G-A	0,47	0,47	0,47	0,47	0,53
HA-G-M	0,44	0,44	0,44	0,44	0,54
HA-G-B	0,40	0,40	0,40	0,40	0,55

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

Quadro 23 – Correlação de ordem entre os valores genéticos reais e preditos, obtidos pela metodologia EBLUP e Bayesiana sob três níveis de informação, em populações com três níveis de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância aditiva e ambiental.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HGA-P-A	0,27	0,23	0,27	0,27
HGA-P-M	FC	0,39	0,40	0,40
HGA-P-B	0,51	0,17	0,51	0,51
HGA-G-A	0,40	0,40	0,40	0,40
HGA-G-M	0,43	0,44	0,44	0,44
HGA-G-B	0,50	0,50	0,50	0,50

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

De forma geral, as correlações foram menores para as populações pequenas, indicando grandes alterações na classificação dos animais. Entretanto, mesmo para as populações de tamanho grande os valores de correlação ficaram sempre abaixo de 0,70. Segundo CREWS e FRANKE (1998), correlações de ordem inferiores a 0,70 podem causar alterações na classificação dos animais, comprometendo a seleção e o progresso genético.

Com exceção da estrutura de população com heterogeneidade genética aditiva e ambiental, os rebanhos de maior variabilidade apresentaram as maiores correlações, sendo que a introdução de maiores níveis de informação não proporcionou diferenças nas correlações. Para as populações com estrutura de heterogeneidade ambiental (Quadro 22), as maiores correlações de ordem foram obtidas para a população grande, quando a distribuição NC foi utilizada sendo de 0,53, 0,54 e 0,55, enquanto que o método EBLUP obteve 0,47, 0,44 e 0,40 nos níveis de alta, baixa e média variabilidade, respectivamente.

Nos Quadros 24 a 28 são apresentadas as correlações de ordem, quando as análises foram realizadas dentro das subclasses de variâncias. De forma geral, as estimativas de correlação de ordem foram baixas para todos os tipos de heterogeneidade estudados, em ambos tamanhos de população. Os métodos REML – BLUP e Bayesiano foram similares na classificação dos indivíduos, com exceção de algumas populações pequenas que tiveram problemas de convergência pelo método REML. Outro fato constatado foi que o aumento no nível de informação que, anteriormente, havia causado uma

melhora considerável na acurácia da estimação dos componentes de variâncias, somente em alguns casos produziu efeito positivo na classificação dos indivíduos. Segundo KENNEDY e HENDERSON (1991), a classificação dos animais, de modo geral, não é afetada devido a erros nas variâncias, quando as análises possuem característica única. Contudo, isto pode não ser verificado para características múltiplas.

As populações com alta variabilidade genética aditiva apresentaram correlações de ordem superiores quando comparadas com populações de baixa variabilidade (Quadro 25). Uma possível explicação para este fato é que, nas populações de maior variabilidade genética, as diferenças entre os valores genéticos dos indivíduos são mais pronunciadas, facilitando aos métodos uma classificação mais acurada dos indivíduos. Entretanto, situação inversa ocorreu com relação à variabilidade ambiental (Quadro 26), onde foi constatado que populações com menor variabilidade apresentaram as maiores correlações de ordem para este grupo, o que era esperado, uma vez que ambientes mais variáveis dificultam o isolamento dos valores genéticos, das demais fontes de variação. Este comportamento foi verificado para ambas as metodologias, sendo que o aumento no nível de informação *a priori* pela metodologia bayesiana não produziu acréscimo considerável na correlação de ordem. Os valores mais altos de correlação de ordem foram observados para as subpopulações grandes, oriundas da população com homogeneidade de variâncias (Quadro 24), independentemente do método empregado ou do nível de informação utilizado, sendo 0,75, 0,78 e 0,75 para os níveis de alta, média e baixa variabilidade, respectivamente.

Para as subpopulações de tamanho pequeno, os valores encontrados foram os mais problemáticos, indicando grandes alterações na classificação dos indivíduos e baixa acurácia na predição dos valores genéticos. Provavelmente, este fato ocorreu porque as subpopulações formadas foram muito pequenas, apresentando desvios na normalidade dos dados.

Quadro 24 – Correlação de ordem entre os valores genéticos reais e preditos, obtidos pelas metodologias EBLUP e Bayesiana, com distribuição normal, sob três níveis de informação e distribuição normal contaminada (NC), com flat *priors*, para as subclasses com iguais níveis de variabilidade e dois tamanhos.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HO-P-1	FC	0,48	0,50	0,50	0,49
HO-P-2	0,60	0,60	0,59	0,59	0,60
HO-P-3	0,69	0,68	0,69	0,68	0,70
HO-P-T	0,59	0,61	0,64	0,66	0,63
HO-G-1	0,75	0,75	0,75	0,75	0,75
HO-G-2	0,78	0,78	0,78	0,78	0,78
HO-G-3	0,75	0,75	0,75	0,75	0,75
HO-G-T	0,78	0,78	0,78	0,78	0,78

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

Quadro 25 – Correlação de ordem entre os valores genéticos reais e preditos, obtidos pelas metodologias EBLUP e Bayesiana, sob três diferentes níveis de informação, para as subclasses de variabilidade genética e dois tamanhos.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HG-P-A	0,69	0,69	0,71	0,70
HG-P-M	0,61	0,60	0,61	0,61
HG-P-B	FC	0,38	0,38	0,38
HG-G-A	0,75	0,75	0,75	0,75
HG-G-M	0,68	0,69	0,69	0,70
HG-G-B	0,58	0,58	0,58	0,59

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

Quadro 26 – Correlação de ordem entre os valores genéticos reais e preditos, obtidos pelas metodologias EBLUP e Bayesiana com distribuição normal, sob três níveis de informação e distribuição normal contaminada (NC) com flat *priors*, para as subclasses de variabilidade ambiental e dois tamanhos.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HA-P-A	0,30	0,25	0,30	0,29	0,26
HA-P-M	FC	0,40	0,41	0,42	0,40
HA-P-B	0,52	0,15	0,52	0,52	0,53
HA-G-A	0,45	0,45	0,45	0,45	0,45
HA-G-M	0,48	0,48	0,48	0,49	0,48
HA-G-B	0,56	0,56	0,56	0,56	0,56

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

Quadro 27 – Correlação de ordem entre os valores genéticos reais e preditos, obtidos pelas metodologias EBLUP e Bayesiana sob três níveis de informação para as subclasses de variabilidade genética e ambiental e dois tamanhos.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HGA-P-A	0,41	0,38	0,40	0,40
HGA-P-M	FC	0,40	0,41	0,41
HGA-P-B	0,36	0,34	0,35	0,35
HGA-G-A	0,52	0,52	0,52	0,52
HGA-G-M	0,49	0,49	0,49	0,49
HGA-G-B	0,45	0,46	0,46	0,46

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

Comparando-se as formas de análises realizadas, pode-se dizer que, em algumas situações, as análises realizadas dentro dos subníveis apresentaram melhores resultados em relação àquelas com as populações totais. Quando a presença de heterogeneidade de variância genética aditiva foi desconsiderada pela metodologia EBLUP (Quadro 21), os valores de correlação de ordem para a população grande foram de 0,68, 0,62 e 0,51, enquanto que as análises dentro dos subníveis (Quadro 25) apresentaram resultados de 0,75, 0,68 e 0,58 para os níveis de alta, média e baixa variabilidade, respectivamente. Esta tendência foi constatada também para as demais estruturas de heterogeneidade, indicando que desconsiderar a presença de heterogeneidade de variâncias pode causar prejuízos para os subníveis de variabilidade formados, sendo preferível, portanto, as análises realizadas isoladamente em cada estratificação de variância.

4.3.2. Quadrado Médio do Erro

No Quadro 28 são apresentados os quadrados médios dos erros (QME), entre os valores genéticos reais e preditos, obtidos pela metodologia EBLUP e Bayesiana com distribuição normal, com três diferentes níveis de informação e distribuição normal contaminada (NC), com flat *priors*, desconsiderando a presença dos diferentes tipos de heterogeneidade.

De forma geral, observou-se que os QME, em ambas as metodologias, foram menores para as populações grandes e maiores para as populações pequenas, indicando menor acurácia das predições para estas populações. Comparando-se as duas metodologias, pode-se observar que os resultados

foram bastante similares. Porém, a utilização de *priors* mais informativos pela metodologia Bayesiana, em todas as situações, conduziu a predição de valores genéticos ligeiramente mais acurados, principalmente para as populações pequenas. Para a estrutura com homogeneidade de variâncias e de tamanho pequeno, os valores obtidos foram de: 137,45, 122,98 e 110,35, para os níveis NI, PI e I, respectivamente. Enquanto os valores obtidos quando a heterogeneidade genética foi desconsiderada foram de: 131,14, 112,91 e 99,66 para os níveis NI, PI e I, respectivamente. Estes resultados demonstram como informações adicionais podem contribuir para o aumento da acurácia na predição dos valores genéticos. Entretanto, observa-se que para as populações de tamanho grande o aumento da acurácia, proporcionado pela utilização de maiores níveis de informação *a priori*, foi pouco expressivo demonstrando a maior importância da informação *a priori* para as populações de tamanho pequeno.

Quadro 28 – QME entre os valores genéticos reais e preditos, obtidos pelas metodologias EBLUP e Bayesiana com distribuição normal, sob três níveis de informação e distribuição normal contaminada (NC) com flat *priors*, desconsiderando a presença de heterogeneidade de variâncias.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HO-P-T	139,93	137,45	122,98	110,35	131,52
HO-G-T	113,19	113,40	113,33	112,32	113,30
HG-P-T	133,01	131,14	112,91	99,66	-
HG-G-T	100,40	99,74	99,74	99,74	-
HA-P-T	232,27	239,27	224,05	223,64	242,85
HA-G-T	183,04	182,11	182,21	182,13	182,57
HGA-P-T	155,50	146,32	134,05	129,82	-
HGA-G-T	154,40	152,62	152,82	151,97	-

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

Os maiores valores encontrados para o QME foram observados quando a heterogeneidade ambiental foi desconsiderada para a população pequena, sendo de: 239,27, 224,05 e 223,64 para os níveis NI, PI e I, respectivamente. Para a população de tamanho grande estes valores oscilaram em torno de 182 em todos os níveis de informação. Pode ser observado também que a utilização da distribuição NC não produziu resultados diferentes para este caso. Na situação onde a HGA foi desconsiderada os valores de QME também foram

altos, porém menores que na HA. Desta forma, fica evidente que desconsiderar a presença de heterogeneidade pode acarretar na predição de valores genéticos menos acurados, principalmente se a heterogeneidade estiver presente no componente ambiental, uma vez que para esta situação, nem mesmo a utilização de *priors* informativos foi capaz de diminuir os erros de predição. Possivelmente, quando a HA é desconsiderada, fatores de perturbação podem ocorrer prejudicando o isolamento dos valores genéticos e conseqüentemente a predição acurada em cada nível de variabilidade.

Segundo MARTINS (2002), a avaliação genética de animais na presença de heterogeneidade entre os rebanhos fica comprometida devido à ponderação dos dados por herdabilidades que não correspondem aos valores reais de cada classe de variabilidade.

Para verificar o efeito da desconsideração da heterogeneidade dentro dos níveis de variabilidade, os resultados obtidos a partir das populações totais foram comparados com os valores genéticos reais dos subníveis formados.

Nos Quadros 29 a 31 são apresentados os QME quando a presença da heterogeneidade de variâncias foi desconsiderada, sendo as análises realizadas com a população total e os resultados comparados com os valores genéticos reais dos subníveis. Pode-se observar que as metodologias apresentaram resultados semelhantes quando *priors* não informativos foram utilizados. As populações pequenas apresentaram os piores resultados para todas as estruturas estudadas e o aumento do nível de informação provocou leve melhora na acurácia das predições, concordando com os resultados obtidos anteriormente. Para todas as estruturas de população com heterogeneidade de variâncias, os valores de QME foram diferentes entre os subníveis. Houve uma tendência dos ambientes de maior variabilidade apresentarem valores mais altos de QME, enquanto os ambiente menos variáveis apresentaram valores mais baixos.

Quadro 29 – QME entre os valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três níveis de informação para as subclasses de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância genética aditiva.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HG-P-A	162,27	159,59	136,42	117,84
HG-P-M	127,62	124,95	106,53	91,73
HG-P-B	111,78	109,68	102,68	99,40
HG-G-A	126,75	125,35	125,23	124,81
HG-G-M	89,99	89,77	89,53	89,41
HG-G-B	84,33	84,35	84,72	84,86

NI= Não informativo; PI = Pouco informativo; I = Informativo.

Quadro 30 – QME entre os valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana com distribuição normal, sob três níveis de informação e distribuição normal contaminada (NC) com *flat priors*, para as subclasses de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância ambiental.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HA-P-A	262,05	284,02	241,87	238,49	291,59
HA-P-M	243,52	245,26	234,69	235,47	248,86
HA-P-B	192,42	188,45	195,58	196,97	188,11
HA-G-A	194,75	195,74	195,49	195,67	194,68
HA-G-M	186,83	185,46	185,64	185,50	187,01
HA-G-B	167,05	164,63	165,01	164,74	166,03

NI= Não informativo; PI = Pouco informativo; I = Informativo.

Quadro 31 - QME entre os valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três níveis de informação, nas subclasses de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância genética aditiva e ambiental.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HGA-P-A	242,75	228,02	205,75	197,15
HGA-P-M	127,38	118,88	108,62	105,28
HGA-P-B	91,93	87,44	82,80	81,93
HGA-G-A	229,94	226,12	226,57	224,49
HGA-G-M	134,30	132,82	132,99	132,33
HGA-G-B	98,61	98,59	98,56	98,76

NI= Não informativo; PI = Pouco informativo; I = Informativo.

Para a situação com HG (Quadro 29), os valores obtidos a partir da metodologia Bayesiana, para a população grande, foram 125,35, 89,77 e 84,35 para os níveis de alta, média e baixa variabilidade, respectivamente. Quando a HA foi desconsiderada (Quadro 30), os valores de QME foram mais altos sendo de 195,74, 185,46 e 164,63, para os níveis de alta, média e baixa variabilidade, respectivamente. Contudo, quando os dois tipos de heterogeneidade foram desconsiderados (Quadro 31), o ambiente mais variável conduziu a um valor de QME de 229,94, sendo o maior encontrado para as populações de tamanho grande. Estes resultados indicam que animais oriundos de rebanhos mais variáveis tendem a ter seus valores genéticos preditos menos acurados, principalmente se a variabilidade estiver associada a fatores ambientais, concordando com as observações de GARRICK e VAN VLECK (1987) que concluem que assumir a homogeneidade, quando existe heterogeneidade, reduz a eficiência da seleção, sendo o ambiente mais variável menos acurado.

Não foram observadas diferenças expressivas entre os resultados obtidos com a distribuição NC e os obtidos com a distribuição normal pela metodologia Bayesiana.

Nos Quadros 32 a 35 são apresentados os QME quando as análises foram realizadas dentro de cada nível de variabilidade. Pode-se observar que as duas metodologias foram semelhantes entre si, com algumas exceções. Por exemplo, quando a análise foi realizada com a população HO-P-3 (Quadro 32), a metodologia EBLUP obteve um QME próximo de 100 e a metodologia Bayesiana próximo de 168. Entretanto, não foi verificado outro resultado como este. Uma das vantagens apresentadas pela metodologia Bayesiana, é que mesmo nas situações onde a metodologia REML não obteve convergência, o QME, obtido com esta metodologia, foi similar aos das demais subpopulações do grupo estudado e o nível de informação *a priori* conduziu a uma melhora considerável na acurácia da predição.

Quadro 32 – QME entre os valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana com distribuição normal, sob três níveis de informação e distribuição normal contaminada (NC) com flat *priors*, para as subclasses de igual variabilidade e dois tamanhos.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HO-P-1	FC	167,63	126,82	118,11	130,41
HO-P-2	144,96	139,55	127,44	128,14	147,19
HO-P-3	100,79	168,84	100,96	101,28	103,18
HO-P-T	139,93	137,45	122,98	110,35	131,52
HO-G-1	126,51	125,27	125,45	124,59	125,86
HO-G-2	113,87	113,36	113,40	113,23	113,72
HO-G-3	128,44	129,58	127,41	125,42	128,09
HO-G-T	113,19	113,40	113,33	112,32	113,30

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

Quadro 33 – QME entre os valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três níveis de informação, para as subclasses de variabilidade genética e dois tamanhos.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HG-P-A	188,76	202,12	109,66	109,88
HG-P-M	106,47	102,85	105,37	106,12
HG-P-B	FC	114,55	105,35	103,88
HG-G-A	126,53	125,52	125,42	124,65
HG-G-M	123,33	117,02	115,00	102,83
HG-G-B	95,06	92,72	92,76	91,16

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

Quadro 34 – QME entre os valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana com distribuição normal, sob três níveis de informação e distribuição normal contaminada (NC) com flat *priors*, para as subclasses de variabilidade ambiental e dois tamanhos.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HA-P-A	240,21	333,46	241,82	241,29	357,24
HA-P-M	FC	231,80	213,60	211,80	214,84
HA-P-B	213,92	263,23	193,31	190,66	190,03
HA-G-A	216,19	219,80	213,96	213,84	214,19
HA-G-M	226,49	216,01	210,59	202,01	218,30
HA-G-B	162,67	163,59	161,86	161,85	161,95

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

Quadro 35 – QME entre os valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três níveis de informação, para as subclasses de variabilidade genética e ambiental e dois tamanhos.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HGA-P-A	233,91	222,70	214,54	213,67
HGA-P-M	FC	126,85	116,67	113,78
HGA-P-B	82,72	84,18	82,23	82,43
HGA-G-A	248,09	243,39	244,15	242,33
HGA-G-M	148,55	145,98	146,46	146,19
HGA-G-B	100,43	97,06	97,12	96,13

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

O nível de informação empregado na metodologia Bayesiana, principalmente nas populações pequenas, proporcionou uma queda no QME em todas as análises realizadas. Apesar desta melhora na acurácia da predição dos valores genéticos não refletir-se na classificação dos animais, ela poderia ter sido mais importante em populações cujos animais apresentassem méritos genéticos mais próximos entre si, de modo que, uma pequena melhora na acurácia fosse suficiente para alterar a classificação dos indivíduos. Outro fator de destaque é que a melhor acurácia dos valores genéticos pode resultar em uma melhor predição dos ganhos genéticos obtidos.

A utilização da distribuição NC (Quadros 32 e 34), utilizada para a estrutura com homogeneidade e com heterogeneidade ambiental, apresentou resultados semelhantes aos obtidos quando a distribuição normal Gaussiana foi empregada. Entretanto, esperava-se que a NC fosse melhor nestas situações por ser uma distribuição mais robusta. Estes resultados indicam que o número de indivíduos contaminantes destas populações provavelmente não é o fator limitante destas análises.

Pode-se observar que as análises realizadas dentro dos subníveis de variabilidade (Quadros 32, 33, 34 e 35), conduziram a maiores valores de QME quando comparadas com os QME resultantes da desconsideração da heterogeneidade (Quadros 29, 30 e 31). Entretanto, estes resultados foram diferentes dos encontrados para as correlações de ordem, onde em geral, os valores obtidos foram maiores para as análises realizadas dentro dos subníveis. Uma provável explicação para estes resultados é que os métodos devem estar conduzindo a erros de proporções semelhantes para todos os

indivíduos, não prejudicando a classificação, apesar de provocar aumentos no QME.

Para a estrutura com heterogeneidade de variância genética aditiva (Quadro 33), a subpopulação pequena, com alta variância genética aditiva, apresentou valor de QME igual a 188,76, sendo este o maior valor encontrado, para esta estrutura. Por outro lado, apresentou correlação de ordem igual a 0,69, sendo a maior dentre as subpopulações pequenas (Quadro 25), indicando que apesar desta situação de alta variabilidade ter resultado em uma predição menos acurada, conduziu a uma classificação mais correta.

Dessa forma, preferem-se os resultados obtidos em relação às Correlações de Ordem, onde as análises realizadas dentro das subclasses foram indicadas como sendo as mais adequadas.

4.3.3. Porcentagens de Indivíduos Selecionados em Comum

Com o objetivo de verificar o efeito dos diferentes tipos de heterogeneidade, níveis de informação e tamanhos de população sobre a seleção, calcularam-se as porcentagens de indivíduos comuns para 15% dos melhores indivíduos oriundos das análises realizadas com as populações totais (Quadro 36).

A porcentagem de indivíduos em comum, obtidos para a estrutura HO-G-T, foi de 62% para ambas as metodologias, sendo o maior valor encontrado. O aumento do nível de informação bem como a utilização da distribuição normal contaminada não provocou mudanças neste resultado. Contudo, o valor obtido para a estrutura HO-P-T foi de 42, 44 e 47% para os níveis NI, PI e I, respectivamente, sendo levemente melhor para o nível mais informativo. Apesar do aumento do nível de informação não ter melhorado substancialmente a porcentagem de indivíduos em comum, os níveis mais informativos apresentaram valores iguais ou superiores, indicando que uma pequena melhora pode ocorrer quando maiores níveis de informação são utilizados pela metodologia Bayesiana.

Quadro 36 – Porcentagem de indivíduos em comum entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três níveis de informação com distribuição normal e normal contaminada (NC) com flat *priors*, desconsiderando a presença de heterogeneidade de variâncias.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HO-P-T	42	42	44	47	44
HO-G-T	62	62	62	62	62
HG-P-T	44	47	47	47	-
HG-G-T	56	56	56	56	-
HA-P-T	38	33	40	40	36
HA-G-T	42	42	42	42	42
HGA-P-T	38	40	40	40	-
HGA-G-T	42	42	42	42	-

Para a estrutura HG-G-T o valor encontrado foi de 56% para ambas as metodologias e níveis de informação, demonstrando uma queda da porcentagem em relação à situação com homogeneidade de variâncias.

Além disso, os valores mais baixos foram encontrados para a estrutura com heterogeneidade ambiental, sendo de 42% para a população grande e de 33 a 40% para a população pequena. Estes resultados confirmam as observações anteriores onde a desconsideração deste tipo de heterogeneidade apresentou os piores resultados tanto para a correlação de ordem quanto para o QME.

Portanto, conclui-se que desconsiderar a presença de heterogeneidade de variâncias, principalmente se for heterogeneidade ambiental pode acarretar em prejuízos no processo de seleção dos melhores animais e conseqüentemente no progresso genético da população.

Nos Quadros 37 a 39 são apresentadas as porcentagens de indivíduos em comum, entre os 15 % melhores valores genéticos reais e preditos, obtidos pela metodologia EBLUP e Bayesiana, quando as análises foram realizadas com a população total e os resultados comparados com os valores genéticos reais dos subníveis de variabilidade.

Neste caso as análises para porcentagens de indivíduos selecionados em comum foram realizadas apenas para *priors* não informativos, uma vez que o objetivo agora é verificar o efeito da heterogeneidade no nível de variabilidade.

Quadro 37 – Porcentagem de indivíduos em comum, entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob flat *priors*, para as subclasses de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância genética aditiva.

Populações	EBLUP	Bayesiana -NI
HG-P-A	27	33
HG-P-M	47	47
HG-P-B	27	33
HG-G-A	64	64
HG-G-M	55	55
HG-G-B	48	48

NI= Não informativo

Quadro 38 – Porcentagem de indivíduos em comum, entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob flat *priors* com distribuição normal (NI-Gauss) e normal contaminada (NI-NC), para as subclasses de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância ambiental.

Populações	EBLUP	Bayesiana	
		NI-Gauss	NI-NC
HA-P-A	40	27	27
HA-P-M	33	40	40
HA-P-B	40	40	40
HA-G-A	41	41	40
HA-G-M	43	42	42
HA-G-B	42	43	43

NI= Não informativo

Quadro 39 – Porcentagem de indivíduos em comum, entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob flat *priors*, para as subclasses com diferentes níveis de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância genética aditiva e ambiental.

Populações	EBLUP	Bayesiana -NI
HGA-P-A	37	48
HGA-P-M	39	39
HGA-P-B	37	38
HGA-G-A	37	38
HGA G-M	47	47
HGA-G-B	27	40

NI= Não informativo

De forma geral, as populações de tamanho grande apresentaram os valores mais altos de porcentagem. Para a estrutura de dados com heterogeneidade genética aditiva, as porcentagens obtidas foram 27, 47, e 27% para população pequena e 64, 55 e 48% para a população grande, (Quadro 37) nos níveis de alta, média e baixa variabilidade, respectivamente. Estes valores demonstram a influência negativa do tamanho pequeno das populações no processo de seleção, à semelhança dos resultados obtidos anteriormente, confirmando, portanto, a maior dificuldade dos métodos em selecionar os melhores indivíduos para estas circunstâncias. Além disso, o ambiente de alta variabilidade genética, à semelhança dos resultados obtidos para a correlação de ordem, apresentou o valor mais alto de porcentagem, mostrando que esta condição favorece a um melhor ordenamento dos indivíduos e conseqüentemente maior acurácia na seleção dos indivíduos.

Para as populações com estruturas de HA e de HGA os resultados obtidos foram mais baixos, oscilando em torno dos 40% (Quadros 38 e 39), demonstrando assim, que a presença da heterogeneidade ambiental dificulta a seleção dos melhores indivíduos. O emprego da distribuição normal contaminada não proporcionou diferenças nos resultados.

Com relação às metodologias empregadas, na maioria das situações, os resultados obtidos foram semelhantes, com algumas exceções. Por exemplo, na estrutura com HGA a metodologia Bayesiana apresentou resultados levemente melhores (Quadro 39).

Nos Quadros 40 a 43 são apresentadas as porcentagens de indivíduos em comum entre os 15 % melhores valores genéticos reais e preditos obtidos pela metodologia EBLUP e Bayesiana, quando as análises foram realizadas dentro dos subníveis de variabilidade.

Pode-se verificar que os resultados, obtidos para as populações de tamanho grande, continuaram sendo maiores que os das populações pequenas e que as duas metodologias avaliadas foram semelhantes entre si. Além disso, o aumento no nível de informação *a priori* não produziu melhores resultados, indicando que o aumento da acurácia, constatada para os componentes de variância, não foi refletida na mesma proporção para a seleção dos melhores indivíduos.

Quadro 40 – Porcentagem de indivíduos em comum, entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três níveis de informação com distribuição normal e normal contaminada, para as subclasses de igual variabilidade e dois tamanhos.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HO-P-1	FC	60	53	60	60
HO-P-2	47	47	47	47	47
HO-P-3	53	60	53	53	53
HO-G-1	61	62	62	62	61
HO-G-2	59	60	60	60	60
HO-G-3	58	58	58	58	58

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

Quadro 41 – Porcentagem de indivíduos em comum, entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três diferentes níveis de informação, para as subclasses de variabilidade genética aditiva e dois tamanhos.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HG-P-A	53	53	53	53
HG-P-M	40	40	40	40
HG-P-B	FC	27	27	27
HG-G-A	69	62	62	62
HG-G-M	49	49	48	49
HG-G-B	42	43	43	44

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

Quadro 42 – Porcentagem de indivíduos em comum, entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana com distribuição normal sob três diferentes níveis de informação e distribuição normal contaminada (NC), para as subclasses de variabilidade ambiental e dois tamanhos.

Populações	EBLUP	Bayesiana			
		NI	PI	I	NC
HA-P-A	33	20	40	48	27
HA-P-M	FC	27	20	27	27
HA-P-B	33	20	33	33	33
HA-G-A	37	37	35	35	35
HA-G-M	41	40	40	39	40
HA-G-B	46	46	45	45	45

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

Quadro 43 – Porcentagem de indivíduos em comum, entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob três diferentes níveis de informação, para as subclasses de variabilidade genética e ambiental e dois tamanhos.

Populações	EBLUP	Bayesiana		
		NI	PI	I
HGA-P-A	27	27	27	20
HGA-P-M	FC	27	27	27
HGA-P-B	40	40	40	40
HGA-P-T	38	40	40	40
HGA-G-A	42	42	42	42
HGA-G-M	38	37	37	37
HGA-G-B	35	36	35	36
HGA-G-T	42	42	42	42

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

Para a estrutura de dados com HG (Quadro 41), os ambientes mais variáveis apresentaram maiores porcentagens de indivíduos em comum, sendo de 62, 49 e 43% para os subníveis de alta, média e baixa variabilidade, respectivamente. Entretanto, situação inversa ocorreu para a estrutura de dados com HA, onde os valores encontrados foram 37, 40 e 46%, confirmando as observações anteriores.

WINKELMAN e SCHAEFFER (1988) destacam que sob seleção intensa e variâncias heterogêneas entre rebanhos, a escolha de touros e vacas poderia resultar na seleção de animais que não são geneticamente os melhores. Segundo esses autores, quando a heterogeneidade é causada por fatores ambientais, a avaliação genética dos animais poderia estar refletindo mais o ambiente do que o seu potencial genético.

Um fato constatado para os ambientes de média variabilidade é que ocorreu uma diminuição da porcentagem de indivíduos em comum, quando as análises foram realizadas dentro dos subníveis, indicando que ao desconsiderar a presença da heterogeneidade, o subnível médio de variabilidade é beneficiado com o aumento de volume dos dados, em detrimento aos subníveis de alta e baixa. Este fato pode ser explicado porque o subnível de média variabilidade possui valor médio de variância genética e ambiental em relação à população total, portanto, beneficiada quando a heterogeneidade é desprezada.

4.3.4. Porcentagens de Indivíduos Selecionados por Ambiente

Para verificar o efeito da desconsideração dos diferentes tipos de heterogeneidade, na seleção dos melhores indivíduos por ambiente, procedeu-se o cálculo da porcentagem de indivíduos selecionados por ambiente entre os 15% melhores valores genéticos reais e preditos.

Nos Quadros 44 a 47 são apresentadas as porcentagens de indivíduos selecionados por ambiente entre os 15% melhores valores genéticos reais e valores genéticos preditos, obtidos pelas metodologias EBLUP e Bayesiana sob flat *priors*, considerando distribuição Gaussiana (NI-GAUSS) e distribuição normal contaminada (NI-NC).

Como as médias entre os subníveis foram iguais, os ambientes com maior variabilidade genética aditiva possuíam uma maior porcentagem de animais superiores. Para população com HG (Quadro 45), entre os 15% melhores valores genéticos reais, 41, 34 e 25% são oriundos dos rebanhos de alta, média e baixa variabilidade, respectivamente. Entretanto, para a população com HO, os três subníveis de variabilidade possuem indivíduos com mérito genético semelhantes, sendo os valores para os três ambientes de 33% (Quadro 44). Para a população com HA (Quadro 46), o comportamento é semelhante, cada subnível possui o mesmo mérito genético, sendo as porcentagens observadas, para a situação real de 34, 34 e 32% para população com alta, média e baixa variabilidade ambiental, respectivamente.

Pode-se observar que na melhor situação analisada, população grande com homogeneidade de variâncias (Quadro 44), as porcentagens de indivíduos selecionados por ambiente, pelos métodos, foram uniformes nas três subpopulações, apresentado valores de 33%. Entretanto, mesmo com homogeneidade de variâncias, as populações pequenas apresentaram resultados não satisfatórios, fato também observado nas demais estruturas estudadas. Portanto, pode-se concluir que o maior problema na estimação dos componentes de variância e predição dos valores genéticos, para estas populações, foi devido ao seu pequeno tamanho e não necessariamente ao efeito da heterogeneidade.

Quadro 44 – Porcentagem de indivíduos selecionados por ambiente entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob flat *priors* (NI), distribuição normal (Gaussiana) e normal contaminada (NI-NC), em populações com iguais níveis de variabilidade e dois tamanhos.

Populações	REAL	EBLUP	Bayesiana	
			NI-Gauss	NI-NC
HO-P-1	22	9	9	9
HO-P-2	46	60	60	60
HO-P-3	31	31	31	31
HO-G-1	33	33	33	33
HO-G-2	33	33	33	33
HO-G-3	33	33	33	33

NI= Não informativo

Quadro 45 – Porcentagem de indivíduos selecionados por ambiente entre os 15% melhores valores genéticos reais e os preditos pela, metodologia, EBLUP e Bayesiana com flat *priors*, em populações com diferentes níveis de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância genética aditiva.

Populações	REAL	EBLUP	Bayesiana -NI
HG-P-A	33	49	44
HG-P-M	53	51	53
HG-P-B	13	0	2
HG-G-A	41	38	38
HG-G-M	34	34	34
HG-G-B	25	28	28

NI= Não informativo

Quadro 46 – Porcentagem de indivíduos selecionados por ambiente entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob flat *priors*, distribuição normal (NI-Gauss) e normal contaminada (NI-NC), em populações com diferentes níveis de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância ambiental.

Populações	REAL	EBLUP	Bayesiana	
			NI-Gauss	NI-NC
HA-P-A	31	51	51	51
HA-P-M	31	27	24	24
HA-P-B	38	22	25	25
HA-G-A	34	38	35	35
HA-G-M	34	34	33	33
HA-G-B	32	28	32	32

NI= Não informativo

Quadro 47 – Porcentagem de indivíduos selecionados por ambiente entre os 15% melhores valores genéticos reais e os preditos pelas metodologias EBLUP e Bayesiana sob flat *priors*, em populações com diferentes níveis de variabilidade e dois tamanhos, desconsiderando a presença de heterogeneidade de variância genética aditiva e ambiental.

Populações	REAL	EBLUP	Bayesiana -NI
HGA-P-A	47	38	44
HGA-P-M	29	38	36
HGA-P-B	24	24	20
HGA-G-A	43	39	40
HGA G-M	38	35	34
HGA-G-B	19	26	25

NI= Não informativo

Na situação onde a HG aditiva foi desconsiderada (Quadro 45), a porcentagem de indivíduos selecionados do ambiente com alta variabilidade foi reduzida de 41%, na situação real, para 38% e para o ambiente com baixa variabilidade aumentou de 25% para 28%, em ambas as metodologias. Portanto, esse fato indica que animais oriundos de subpopulações com maior variabilidade genética aditiva, quando as médias são iguais e a HG é desconsiderada, são subestimados e os animais oriundos de subpopulações com baixa variabilidade são superestimados. Segundo MARTINS (2002), animais criados em ambientes com maior variância genética aditiva tem seus dados subvalorizados, uma vez que os dados corrigidos para os efeitos fixos de ambiente têm em sua composição, proporcionalmente, maior contribuição genética do que a ponderação que lhes estaria sendo aplicado, enquanto animais criados em ambientes de menor variância podem ser selecionados equivocadamente.

Para a população onde a HA foi desconsiderada, os resultados obtidos, pelo método EBLUP, para população grande, foram 38, 34 e 28% (Quadro 46) para as subpopulações com alta, média e baixa variabilidade, respectivamente. Estes resultados, quando comparados com os valores reais, demonstram que se a HA é desconsiderada, animais oriundos de ambientes mais variáveis tendem a ter os seus valores genéticos superestimados, diferentemente da HG, visto que os dados corrigidos para os efeitos fixos de ambiente têm em sua composição, proporcionalmente, menor contribuição genética do que a ponderação que lhes estaria sendo aplicada. Resultados semelhantes foram

obtidos por POWER et al. (1983); HILL et al. (1983); BOLDMAN e FREEMAN (1990); TORRES (1998) e GARRICK (1989) que verificaram a existência de uma tendência de animais de ambientes mais variáveis serem selecionados.

Dessa forma, animais criados ou então com maior número de filhos em ambientes mais variáveis poderiam ser equivocadamente selecionados, concordando com as observações de MARTINS (2002). Para esta situação a metodologia Bayesiana apresentou resultados mais próximos da situação real, 35, 33 e 32 para população com alta, média e baixa variabilidade, respectivamente.

Na situação com HGA (Quadro 47) os resultados para os níveis de alta, média e baixa variabilidade foram respectivamente 39, 35 e 26%, sendo os valores reais de 43, 38 e 19%. Portanto, houve uma tendência de superestimar os animais oriundos da subpopulação com baixa variabilidade e subestimar os oriundos da subpopulação com alta variabilidade. EVERETT et al. (1982), ao trabalharem com dados reais, encontraram que em rebanhos de iguais méritos genéticos, ambientes com maior variância fenotípica tem uma maior proporção de vacas de elite identificadas do que rebanhos com menor variância.

HILL (1984) e VISON (1987) também relatam que existe uma tendência de indivíduos de grupos mais variáveis serem selecionados, e isto pode estar associado à redução na acurácia da predição dos valores genéticos, conseqüentemente com redução no ganho genético. POWER et al. (1983) também concluíram que uma maior proporção de indivíduos poderia ser selecionada de rebanhos com maior herdabilidade.

5. CONCLUSÕES

O aumento do nível de informação *a priori* conduz a melhores estimativas de componentes de variância no método Bayesiano, principalmente para populações pequenas. Entretanto, quando *priors* não informativos são utilizados, o método REML e o método Bayesiano apresentam resultados semelhantes.

A desconsideração dos diferentes tipos de heterogeneidade de variâncias causa problemas nas estimativas dos componentes de variâncias. Nos subníveis de alta e baixa variabilidade, as estimativas são subestimadas e superestimadas, respectivamente, sendo que a subpopulação com variabilidade média é menos prejudicada.

A predição dos valores genéticos dos animais é mais acurada quando maiores níveis de informação são utilizados. Porém, esta maior acurácia não é refletida, na mesma proporção, para a classificação dos animais.

A presença da heterogeneidade de variâncias causa problemas para a seleção dos melhores indivíduos, principalmente se a heterogeneidade estiver presente no componente ambiental, sendo os animais selecionados equivocadamente do ambiente mais variável.

O método Bayesiano é preferido para populações de tamanho pequeno quando há disponibilidade de *priors* informativos. A utilização da distribuição normal contaminada não é eficiente em eliminar problemas causados pela heterogeneidade de variâncias. Entretanto, produz resultados semelhantes ou superiores aos encontrados pela distribuição normal Gaussiana.

REFERÊNCIAS BIBLIOGRÁFICAS

- BALIEIRO , J.C.C., LOPES, P.S., ELER, J.P. et al. Efeito da heterogeneidade de variância na avaliação genética de bovinos da raça Nelore: Análises de características múltiplas para peso à desmama. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 39, Recife, PE. **Anais...** Sociedade Brasileira de Zootecnia. CD-ROM. Melhoramento genético animal. 2002 (a).
- BALIEIRO, J.C.C. **Heterogeneidade de variância na avaliação genética de bovinos da raça Nelore**. Viçosa: Universidade Federal de Viçosa, 2001. 88 p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 2001.
- BALIEIRO, J.C.C., LOPES, P.S., ELER, J.P. et al. Efeito da heterogeneidade de variância na avaliação genética de bovinos da raça Nelore: Análises de característica única para peso à desmama em escala original e transformada. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 39, Recife, PE. **Anais...** Sociedade Brasileira de Zootecnia. CD-ROM. Melhoramento genético animal. 2002 (b).
- BOLDMAN, K. G., FREEMAN, A. E., Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. **Journal of Dairy Science**, v. 73, p. 503-512, 1990.
- BOLDMAN, K. G., KRIESE, L. A., VAN VLECK, L. D., VAN TASSEL, C. P., KACHMAN, S. D. **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, 120 p.1995.
- BROTHERSTONE, S., HILL, W. G. Heterogeneity of variance amongst herds for milk production. **Animal Production**, v. 42, n. 3, p. 297-303. 1986.

- CAMPÊLO, J. E. G. **Heterogeneidade de variância na avaliação genética de bovinos da raça tabapuã.** , 2001. 104 p. Tese (Doutorado em Zootecnia) – Universidade Federal de Viçosa, 2001.
- CARDOSO, F.F.; CARDELLINO, R.A.; CAMPOS,L.T. Fatores ambientais sobre escores de avaliação visual à desmama em bezerros Angus criados no Rio Grande do Sul. **Revista Brasileira de Zootecnia.**, v. 30, n. 2 p318-325, 2001.
- CARNEIRO, A. P. S., **Avaliação genética na presença de heterogeneidade entre rebanhos, utilizando dados simulados.** Viçosa, MG: Universidade Federal de Viçosa, 2003. 81 p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2003.
- CARVALHEIRO, R. **Flexibilizando a matriz R na predição de valores genéticos.** Jaboticabal – SP, 2000. 107 p. Tese (Mestrado em Zootecnia) – Universidade Estadual Paulista, 2000.
- CREWS Jr., D. H. FRANKE, D. E. Heterogeneity of variances for carcass traits by percentage brahman inheritance. **Journal of Animal Science**, v. 76, n. 7, p. 1803-1809. 1998.
- 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.
- EVERETT, R. W., KEOWN, J. F., TAYLOR, J. F. The problem of heterogeneous within herd variances when identifying elite cows. **Journal of Dairy Science**, v. 65, p. 100, (Suppl. 1), (Abstract). 1982.
- FAMULA, T. R. Detection of heterogeneous variance in herd production groups. **Journal of Dairy Science**, v. 72, n.3, p.715-721, 1990.
- FRIDRICH, A. B.; SILVA, M. A.; VENTURA, R.V.; CORRÊA, G. S. S.C.; FERREIRA, I. C.; RUFINO, J. E.; EUCLYDES, R. F.; VALENTE, B. D.; REIS, G. L. Influência da heterogeneidade de variância na classificação de

aves utilizando dados simulados. **V Simpósio da Sociedade Brasileira de Melhoramento Animal**. Pirassununga – SP. 2004.

GARRICK, D. J., POLLAK, R. L., QUAAS, R. L., VAN VLECK, L. D. Variance heterogeneity in direct and maternal weight traits by sex percent purebred for Simmental-sired calves. **Journal of Animal Science**, v. 67, n. 10, p. 2515-2528, 1989.

GARRRICK, D.J.; VAN VLECK, L.D. Aspects of selection for performance in several environments with heterogeneous variances. **Journal of Animal Science**. V. 65, n.2, p. 409-421, 1987.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculation marginal densities. **Journal of the American Statistical Association**, 87, 523-532, 1990.

GEMAN, S., GEMAN, D. **Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images**. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6: 721-741, 1984.

GIANOLA, D., FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, 63: 217-244, 1986.

GIANOLA, D., FOULLEY, J.L., FERNANDO, R.L. et al. Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. **Journal of Dairy Science**, v. 75, n. 10, p. 2805-2823, 1992.

HILL, W. G. On selection among groups with heterogeneous variance. **Animal Production**, v. 39, n. 3, p. 473-477, 1984.

HILL, W. G., EDGWARD, M.R., AHMED, M.K.A. et al. Heritability of milk yield and composition at different levels and variability of production. **Animal Production**, v. 36, n. 1, p. 59-68, 1983.

JENSEN, J.; MADSEN, P., Bayesian estimation of within and across country genetic parameters for MACE. **Danish Institute of Agricultural Sciences**,

Dept of Animal Breeding and Genetics, Research Centre Foulum, P.O. box 50, DK – 8830 Tjele Kenmark, 1992.

KACHAMN, S. D., EVERETT, R. W. A Multiplicative Mixed Model When the Variances Are Heterogeneous. **Journal of Dairy Science**, v. 76 n.03, p. 859-867, 1993.

KENNEDY, B. W., HENDERSON, C. R. The unfinished legacy. **Journal of Dairy Science**, v.74, n. 11, 4067-4081, 1991.

LARSON, H.J. **Introduction to probability theory and statistical inference**. John Wiley & Sons, Singapore, 3 ed., 1982, 637p.

LEANDRO, R.A. **Introdução à estatística Bayesiana**. In: REUNIÃO ANUAL DA REGIÃO DE BIOMETRIA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46 E SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 9. ESALQ/USP, Piracicaba, SP, 51p. 2001.

LEE, K.J., PARK D.K. Heterogeneity of milk production levels in Korean dairy herds. In: WORLD CONGRESS ON GENETICS APLLIED TO LIVESTOCK PRODUCTION, 5, Ontário. **Proceedings...**, v. 17, p. 46-49. 1994.

LITTELL, R.C.; FREUND, R.J.; SPECTOR, P.C. **SAS[®] System for linear models**. 3.ed. Cary, NC: SAS Institute Inc., 329p.1991.

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/Gmosis. **CD-ROM**. Melhoramento Genético Animal. 2000.

MARTINS, E.N.. Avaliação genética e heterogeneidade de variância. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 39, 2002, Recife, PE. **Anais... Sociedade Brasileira de Zootecnia**. CD-ROM. Melhoramento genético animal. 2002

- MEUWISSEN, T.H.E., JONG, G., ENGEL, B. Joint estimation of breeding values and heterogeneous variances of larges data sets. **Journal of Dairy Science**, v. 79, n. 2, p. 310-316, 1996.
- OLIVEIRA, C. A. L.; MARTINS, E. N.; FREITAS, A.R.; ALENCAR, M. M. Heterogeneidade de variâncias nos grupos genéticos formadores da raça Canchim. **Revista Brasileira de Zootecnia** v.30, n.4 Viçosa 2001.
- OLIVEIRA, H. N.; LÔBO, R. B.; BEZERRA, L. A. F. Análise bayesiana na estimação de componentes de (co)variância e predição de valores genéticos do peso aos 365 dias de bovinos canchim **V Simpósio da Sociedade Brasileira de Melhoramento Animal**. Pirassununga – SP. 2004.
- PEREIRA, I.G., OLIVEIRA, H. N., ROSA G.J.M. Simulation study on linear mixed models with contaminated normal distribution in animal breeding. In: **7th World Congress on Genetics Applied to Livestock Production**. Montpellier - França, 2002.
- POWELL, R. L., NORMAN, H. D., WEILAND, B. T. Cow evaluation at different milk yields of herds. **Journal of Dairy Science**, v. 66, n. 1, p. 148-154, 1983.
- ROSA G.J.M.; Robust mixed linear models in quantitative genetics: bayesian analysis via gibbs sampling. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 39, Recife, PE. **Anais... Sociedade Brasileira de Zootecnia**. CD-ROM. Melhoramento genético animal. 2002.
- ROSA, G. J. M.; PADOVANI, C. R. Modelos lineares mistos robustos utilizando-se a distribuição normal contaminada. **Revista de Matemática e Estatística**, São Paulo, v. 18, p. 65-81, 2000.
- SANTOS, G. B.; OLIVEIRA, H. N.; ROSA, G. J. M.; MARQUES, L. F. A.. Aplicação de modelos mistos na predição de valores genéticos de bovinos da raça simental. **V Simpósio da Sociedade Brasileira de Melhoramento Animal**. Pirassununga – SP. 2004.

- SCHENKEL, F. S. **Studies on effects of parental selection on estimation of genetic parameters and breeding values of metric traits.** University of Ghelph, Ghelph, Canada. (PhD. Thesis). 1998.
- SORENSEN, D.; GIANOLA, D. Likelihood, Bayesian and MCMC methods in quantitative genetics. Springer, USA, 740p. 2002.
- 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, 1980.
- TORRES, R.A. **Efeito da heterogeneidade de variância na avaliação genética de bovinos da raça Holandesa no Brasil.** Belo Horizonte, MG: Universidade Federal de Minas Gerais, 1998. 124 p. Tese (Doutorado em Ciências Animal) - Universidade Federal de Minas Gerais, 1998.
- TORRES, R.A.; BERGMAN J. A. G.; COSTA, C. N.; PEREIRA, C.; VALENTE, J.; PENNA, V. M.; TORRE FILHO, R. A.; ARAÚJO, C. V. heterogeneidade de variância e avaliação genética de bovinos da raça Holandesa no Brasil. **Revista Brasileira de Zootecnia**, v.4, p. 1050 - 1059, 2000.
- VINSON, W. E. Potential bias in genetic evaluations from differences in variation within herds. **Journal of Dairy Science**, v. 70, n. 11, p. 2450-2455, 1987.
- WANG, C. S., GIANOLA, D.; SORENSEN, D. A.; JENSEN, J.; CHRSTENSEN, A.; RUTHLEDGE, J.J. Response to selection for litter size in danish Landrace pigs: a Bayesian analysis. **Theory Applied Genetics**, 88:220-230. 1994.
- WANG, C.S., RUTLEDGE, J.J., GIANOLA, D. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. **Genetics Selection and Evolution**. 25: 41-62, 1993.
- 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.

WELLER, J.I., RON, M., BAR-ANAN, R. Accounting for environmentally dependent variances components in BLUP sire evaluation. **Journal of Dairy Science**, v. 68, n. 1, p. 212-213, 1985.

WIGGANS, G. R., VAN RADEN, P. M. Method and effect of adjustment for heterogeneous variance. **Journal of Dairy Science**, v. 74, n. 12, p. 4350-4357, 1991.

WINKELMAN, A., SCHAEFFER, S. R., Effect of heterogeneity of variance on dairy sire evaluation. **Journal of Dairy Science**, v. 71, p. 3033-3039, 1988.