

GISELLE MARIANO LESSA DE ASSIS

**EFEITO DO NÚMERO DE GENES NA AVALIAÇÃO GENÉTICA
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

A848e
2005 Assis, Giselle Mariano Lessa de, 1973-
Efeito do número de genes na avaliação genética
utilizando dados simulados / Giselle Mariano Lessa de
Assis. – Viçosa : UFV, 2005.
xi, 102f. : il. ; 29cm.

Inclui apêndices.

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

Referências bibliográficas: f. 90-98.

1. Melhoramento animal - Métodos de seleção -
Simulação por computador. 2. Método EBLUP.
3. GENESYS (Programa de computador). 4. Análise
de variância. 5. Teoria bayesiana de decisão estatística.
I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 636.0824

GISELLE MARIANO LESSA DE ASSIS

**EFEITO DO NÚMERO DE GENES NA AVALIAÇÃO GENÉTICA
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^a. Carmen Silva Pereira

Prof. Ricardo da Fonseca

Prof. Paulo Sávio Lopes
(Conselheiro)

Prof. Robledo de Almeida Torres
(Conselheiro)

Prof. Ricardo Frederico Euclides
(Orientador)

Aos meus pais Tania e Fernando;
aos meus irmãos Andréa, Tatiana e Victor;
ao meu marido Júnior;
à minha filha Gabriela,

dedico este trabalho.

AGRADECIMENTOS

À Universidade Federal de Viçosa, em especial ao Departamento de Zootecnia e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade de realização deste curso.

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

Ao meu orientador, Prof. Ricardo Frederico Euclides, pela confiança, pela orientação, pela amizade e por disponibilizar o programa Genesys, viabilizando a execução deste trabalho.

Aos Professores Paulo Sávio Lopes, Robledo de Almeida Torres, Carmen Silva Pereira e Cosme Damião Cruz pelos valiosos ensinamentos durante todo o curso, pelas sugestões e pela amizade.

Aos colegas e amigos de curso, Adriana, Aldrin, Amauri, Claudinho, Elivalto, Elizângela, Fábio, Fabiano, Fausto, Felipe, Fernanda, Gustavo, Jaime, Jane, José Filho, Leandro, Lindenbergh, Luciara, Marcelo, Marcos Vinícius, Paulo Nobre, Paulinho, Policarpo, Rachel, Rodolpho, Samuel, Urbano e Willian, pela amizade, companheirismo e pelo convívio agradável.

Aos amigos Ricardinho e Lú, Cristina e Juninho, Céphora, Danielle, Lana e Renatinha pelo constante incentivo e pelos momentos agradáveis e descontraídos.

À Alice e Margarida pelo empenho e dedicação no dia-a-dia, que muito me ajudaram na elaboração desse trabalho.

À minha querida avó Marília e aos meus pais Fernando e Tania, pela dedicação e por me apoiarem sempre, com muita confiança e carinho.

Aos meus irmãos Andréa, Tatiana e Victor, pelo constante incentivo, pela compreensão e pela amizade.

Ao meu marido Júnior, pelo apoio diário, pela paciência e pelas críticas e sugestões, indispensáveis na elaboração deste trabalho.

À minha pequena filha Gabriela, por renovar diariamente minhas forças, sendo exemplo de esperança e amor para todos nós.

A todos que direta ou indiretamente contribuíram para a realização deste trabalho.

BIOGRAFIA

GISELLE MARIANO LESSA DE ASSIS, filha de Fernando Antonio de Assis e Tania Ludmila de Assis, nasceu no Rio de Janeiro, RJ, em 7 de novembro de 1973.

Em dezembro de 1992, concluiu o curso técnico de Biotecnologia pela Escola Técnica Federal de Química no Rio de Janeiro - RJ.

Em dezembro de 1997, graduou-se em Zootecnia pela Universidade Federal de Viçosa, em Viçosa - MG.

De fevereiro de 1998 a março de 1999, foi bolsista de Aperfeiçoamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) na Embrapa Gado de Leite, na área de Melhoramento Genético de Plantas Forrageiras.

Em 23 de março de 2001, obteve o título de Mestre em Genética e Melhoramento na Universidade Federal de Viçosa, em Viçosa - MG.

Em abril de 2001, iniciou o curso de Doutorado em Genética e Melhoramento na Universidade Federal de Viçosa.

Em 14 de fevereiro de 2005, submeteu-se aos exames finais de defesa de tese para obtenção do título de *Doctor Scientiae* em Genética e Melhoramento.

ÍNDICE

RESUMO	viii
ABSTRACT	x
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1. O Modelo Infinitesimal	3
2.1.1. Simulação Computacional de Sistemas Genéticos	8
2.2. A Metodologia de Modelos Mistos	9
2.2.1. Dados sob seleção	12
2.3. Método da Máxima Verossimilhança Restrita	15
2.3.1. Dados sob seleção	16
2.4. Análise Bayesiana	19
2.4.1. Teorema de Bayes	22
2.4.2. Seleção e Acasalamento Não Aleatório	24
2.4.3. Métodos Bayesianos Aplicados	26
2.4.3.1. Distribuição <i>a posteriori</i> completa	27
2.4.3.2. Amostragem de Gibbs	29
3. MATERIAL E MÉTODOS	32
3.1. Simulação dos dados	32
3.2. Análises dos Valores Reais	34
3.2.1. Curtose, Assimetria e Teste de Normalidade	34
3.2.2. Componentes de Variância	36
3.3. Avaliação Genética	36
3.3.1. Estruturas Consideradas	36
3.3.2. Análise Bayesiana <i>via</i> Amostragem de Gibbs	38
3.3.2.1. Modelo	38

3.3.2.2. Distribuições <i>a priori</i> -----	39
3.3.2.3. Cadeia de Gibbs -----	40
3.3.3. EBLUP -----	41
3.3.4. Critérios de Comparação das Metodologias -----	41
3.3.4.1. Componentes de Variância -----	41
3.3.4.2. Valores Genéticos -----	42
3.3.5. Adequabilidade do Modelo Infinitesimal -----	43
3.3.6. Análise genética em gerações avançadas -----	43
4. RESULTADOS E DISCUSSÃO -----	44
4.1. Análises dos Valores Reais -----	44
4.1.1. Curtose, Assimetria e Teste de Normalidade -----	44
4.1.2. Componentes de Variâncias Reais -----	46
4.2. Estimação dos Componentes de Variância -----	51
4.2.1. Porcentagem de Erro -----	60
4.2.2. Comparação Entre Média e Moda <i>a posteriori</i> -----	66
4.2.3. Adequabilidade do Modelo Infinitesimal -----	68
4.2.4. Estimação em Populações Avançadas -----	73
4.3. Predição dos Valores Genéticos -----	75
4.3.1. Quadrado Médio do Desvio -----	75
4.3.2. Porcentagem de Seleccionados em Comum -----	80
4.3.3. Correlação de Ordem -----	82
4.3.4. Avaliação Considerando Diferentes Populações-Base -----	86
5. CONCLUSÃO -----	88
REFERÊNCIAS BIBLIOGRÁFICAS -----	90
APÊNDICE -----	99

RESUMO

ASSIS, Giselle Mariano Lessa de, D.S. Universidade Federal de Viçosa, fevereiro de 2005. **Efeito do número de genes na avaliação genética utilizando dados simulados.** Orientador: Ricardo Frederico Euclides. Conselheiros: Paulo Sávio Lopes e Robledo de Almeida Torres.

Foram simulados quatro distintos tipos de populações por meio do programa GENESYS com os objetivos de: verificar a influência do número de genes e do tamanho da população na estimação de componentes de variância e na predição de valores genéticos; verificar a adequabilidade do modelo infinitesimal como pressuposição nas análises genéticas; comparar as metodologias clássica e Bayesiana na análise genética de dados selecionados; e verificar a influência do nível de informação *a priori* ao utilizar metodologia Bayesiana. Dois processos seletivos foram aplicados por 10 gerações a partir da população-base formada por 120 (população pequena) ou por 2.400 indivíduos com registros (população grande): Seleção ao Acaso e Seleção Fenotípica. Foi considerado que dois diferentes números de genes governavam a característica sob seleção, para cada tamanho de população: 900 ou 10 pares de locos. Para as populações pequenas, foram realizadas 500 repetições de cada processo seletivo e para as populações grandes, 300 repetições. Na análise Bayesiana, três níveis de informação *a priori* foram considerados: não-informativo, pouco informativo e informativo. Os componentes de variância foram estimados utilizando-se somente a população-base, somente a população da 10^a geração após seleção ou todas as populações, desde a população-base até a 10^a geração após seleção. Os valores genéticos foram preditos para a população-base e para a 10^a geração após seleção, considerando, porém, diferentes conjuntos de dados no processo de predição. A Porcentagem de Erro entre os componentes de variâncias estimados e os reais foi utilizada para comparar as metodologias, assim como as diferentes populações e gerações analisadas. Os valores genéticos, por sua vez, foram comparados por meio do Quadrado Médio do Desvio, da Porcentagem de

Indivíduos Selecionados em Comum entre os 15% melhores indivíduos e pela Correlação de Ordem entre os valores reais e preditos. Conforme os resultados obtidos, pôde-se concluir que quando a característica é governada por elevado número de genes, os componentes de variância genética aditiva e ambiental são satisfatoriamente estimados em populações selecionadas grandes ou pequenas pelas metodologias usuais, desde que os registros de todos os indivíduos e a matriz completa de parentesco sejam conhecidos. Por outro lado, quando a característica é governada por reduzido número de genes, estimativas menos acuradas do componente de variância genética aditiva são obtidas em populações grandes e, caso as informações de parentescos e registros anteriores sejam desconhecidos, o erro na estimação desse componente aumenta consideravelmente, em populações grandes ou pequenas. Verificou-se também que os valores genéticos são superestimados sob seleção fenotípica quando os registros de todos os indivíduos e a matriz completa de parentesco são incluídos nas análises, independentemente do tamanho da população. A queda na acurácia é ainda mais acentuada quando a característica é governada por reduzido número de genes, sendo a classificação correta dos indivíduos também prejudicada. A inclusão do registro de todos os indivíduos, assim como da matriz de parentesco completa beneficiam a classificação adequada dos indivíduos. Verificou-se também que o modelo infinitesimal não é adequado para ser utilizado como pressuposição nas análises genéticas quando a característica é governada por poucos genes, independentemente do tamanho da população. Ao comparar as metodologias REML e Bayesiana verificou-se que, em geral, essas metodologias produzem resultados bastante semelhantes na estimação dos componentes de variância. Para análises com menor quantidade de dados, no entanto, estimativas mais acuradas são obtidas ao se utilizar *priors* informativos por meio da análise Bayesiana. Concluiu-se também que a acurácia na predição dos valores genéticos, assim como a classificação dos indivíduos não são alteradas pelo nível de informação *a priori* das análises Bayesianas, cujos resultados também se assemelham aos da metodologia EBLUP.

ABSTRACT

ASSIS, Giselle Mariano Lessa de, D.S. Universidade Federal de Viçosa, February, 2005. **Effect of number of genes on genetic evaluation using simulated data.** Adviser: Ricardo Frederico Euclides. Committee members: Paulo Sávio Lopes and Robledo de Almeida Torres.

Four different population types were simulated using GENESYS program with the following objectives: to verify the influence of the number of genes and the population size on variance component estimation and on breeding values prediction; to verify the infinitesimal model as an appropriate assumption on genetic analyses; to compare the classic and Bayesian methodologies on the genetic analysis of selected data; and to verify the influence of *a priori* information level in Bayesian methodology. Two selective processes were applied for 10 generations starting from base population formed by 120 (small population) or by 2,400 individuals with records (large population): Random Selection and Phenotypic Selection. It was considered that two different numbers of genes governed the trait under selection, for each population size: 900 or 10 pairs of *loci*. Five hundred repetitions of each selective process for small populations and three hundred repetitions were accomplished for large populations. On Bayesian analysis, three *a priori* information levels were considered: no-informative, slightly informative and informative. Variance components were estimated using only base population, only population of the 10th generation after selection or all of populations, from base population up to 10th selection generation. Breeding values were predicted for base population and for 10th selection generation, considering, however, different groups of data on the prediction process. Error Percentage between estimated and real variance components was used to compare the methodologies, as well as the different populations and generations analyzed. Genetic values were compared using Average Square Deviation, Percentage of Common Individuals selected among the 15% better individuals and Rank Correlation among predicted and real values. According to the results, it was concluded that when the trait is

governed by high number of genes, the genetic additive and environmental variance component are well estimated by usual methodologies in large or small selected populations, since data of all animals and complete relationship matrix are known. On the other hand, when the trait is governed by reduced number of genes, less accurate estimates of additive genetic variance are obtained in large populations and, when relationship information and previous data are unknown, estimate errors of that component increase considerably, in large or small populations. It was also verified that breeding values are overestimated under phenotypic selection when data of all individuals and complete relationship matrix are included on analyses, independently of population size. Accuracy decrease is more accentuated when the trait is governed by reduced number of genes, being the correct classification of individuals also affected. The inclusion of all data, as well as complete relationship matrix benefit the appropriate classification of individuals. It was also verified that the infinitesimal model is not appropriate to be used as assumption in genetic analyses when the trait is governed by few genes, independently of population size. When comparing REML and Bayesian methodologies, it was verified that, in general, these methodologies produce similar results on variance components estimation. However, when analyses are performed with smaller amount of data, informative priors using Bayesian analysis yields more accurate estimates. Finally, accuracy of breeding values prediction, as well as the rank of individuals are not changed by *a priori* information level on Bayesian analyses, whose results are also similar to the EBLUP methodology.

1. INTRODUÇÃO

A predição acurada dos valores genéticos é de grande importância nos programas de melhoramento animal, principalmente em espécies que possuem relativamente poucos indivíduos com alta taxa reprodutiva, uma vez que exercem grande influência no mérito genético da população. Geralmente, os dados disponíveis para predição de valores genéticos e para realização de inferências sobre os parâmetros da população são obtidos a campo, a partir de rebanhos onde os acasalamentos não são realizados ao acaso pelos produtores, os quais selecionam os animais e determinam cruzamentos específicos.

Nas análises genéticas, há o interesse constante por metodologias ou estratégias de análises que produzam estimativas acuradas dos componentes de variância para as diversas situações encontradas na prática. Essas estimativas são utilizadas, principalmente, na predição da resposta à seleção e na predição dos valores genéticos dos animais. Estudos de simulação indicam que estimativas não-viesadas dos componentes de variância podem ser obtidas, caso os registros de todos os animais e a matriz de parentesco completa até a população-base sejam incluídos na análise. Porém, esses estudos normalmente utilizam como pressuposição o modelo infinitesimal, além de não incluírem no modelo outros efeitos fixos, a não ser a média populacional. Apesar de simplificarem o modelo, essas condições se afastam consideravelmente das situações reais.

Estudos de mapeamento e detecção de QTL (locos de características quantitativas) vêm mostrando que existem genes com efeitos expressivos (não mais infinitesimalmente pequenos) que governam características quantitativas (Barton e Keightley, 2002). Assim, as metodologias usualmente utilizadas na avaliação genética que se baseiam no modelo infinitesimal podem não ser as mais adequadas. Experimentos de seleção em longo prazo têm confirmado a

inadequabilidade do modelo infinitesimal. No entanto, na análise de gerações iniciais, este modelo parece ser apropriado.

Apesar da maior compreensão atual sobre os efeitos estatísticos e genéticos dos processos seletivos, não existe ainda uma solução estatística geral para o problema da seleção (Schaeffer et al., 1998).

O método usualmente empregado na obtenção do melhor preditor linear não-viesado (BLUP) por meio das Equações de Modelos Mistos, em que os valores paramétricos dos componentes de variância são substituídos por suas estimativas de máxima verossimilhança restrita (conhecido como *BLUP Empírico* ou *EBLUP*), vem sendo questionado, principalmente quando se sabe que a seleção ocorreu. Assim, inferências baseadas em dados selecionados podem ser incorretas.

Recentemente, métodos Bayesianos começaram a ser utilizados na avaliação genética de animais, permitindo inclusive a análise de dados selecionados (Sorensen et al., 1994; Van Tassell et al., 1995) e necessitam ser avaliados e comparados com os métodos usuais, como o EBLUP (Shenckel, 1998). A Amostragem de Gibbs vem sendo usada como uma importante ferramenta de integração numérica, permitindo ampla aplicação da análise Bayesiana no melhoramento animal (Wang et al., 1994a; Wang et al., 1994b). Assim, esse método permite construir uma descrição completa da resposta à seleção por meio da densidade marginal *a posteriori*, onde não apenas o processo da seleção é levado em consideração, mas também a incerteza sobre os valores dos efeitos fixos e dos componentes de variância.

Dentro do contexto apresentado, os objetivos desse trabalho foram:

- verificar a influência do número de genes e do tamanho da população na estimação dos componentes de variância e na predição dos valores genéticos;
- verificar a adequabilidade do modelo infinitesimal como pressuposição nas análises genéticas;
- comparar as metodologias clássica EBLUP e Bayesiana na análise genética de dados selecionados; e
- verificar a influência do nível de informação *a priori* ao utilizar metodologia Bayesiana na análise genética.

2. REVISÃO DE LITERATURA

2.1. O Modelo Infinitesimal

O modelo genético usualmente empregado na análise de dados no melhoramento animal pressupõe que:

- um número muito grande de locos (estritamente infinito) governa a característica de interesse;
- cada loco contribui com efeito infinitesimalmente pequeno para a variância genética aditiva;
- os locos possuem somente ação aditiva, sendo desconsiderada a existência de dominância e epistasia;
- a segregação ocorre independentemente, ou seja, não há ligação entre os genes.

Essas restrições, em conjunto, caracterizam o chamado *modelo infinitesimal*, introduzido por Fisher (1918).

Uma importante consequência desse modelo genético é que, ocorrendo seleção direcional, as mudanças nas freqüências gênicas são infinitesimalmente pequenas, sendo a alteração na variância genética aditiva causada somente pelo desequilíbrio gamético de ligação entre os locos que afetam a característica (Bulmer, 1971). Apesar das mudanças nas freqüências gênicas serem desprezíveis, elas são capazes de alterar a média genética quando somados esses efeitos infinitesimalmente pequenos. Assim, após seleção e acasalamentos, pode-se ainda considerar que os valores genéticos seguem aproximadamente distribuição normal multivariada.

A contribuição do desequilíbrio é negativa, uma vez que a seleção direcional introduz uma correlação também negativa entre pares de locos. Caso a seleção deixe de ocorrer, essa correlação rapidamente desaparece

conforme o equilíbrio é restabelecido, e a variância retorna ao seu valor original (Bulmer, 1971). Porém, este é um resultado limitado, sendo somente verdadeiro quando o número de locos é infinito.

Caso o modelo infinitesimal não se aplique, mudanças na frequência gênica média ocorrem e a variância genética sofre alterações permanentes. A magnitude de qualquer mudança permanente decresce conforme o número de locos envolvidos aumenta, além de que, quando o número de locos é grande, provavelmente essa mudança é bem menor do que a mudança temporária causada pelo desequilíbrio gamético de ligação.

Em populações de tamanho finito, submetidas à seleção, a variância genética aditiva diminui não somente pelo desequilíbrio gamético, mas também devido à endogamia, considerando o modelo infinitesimal (Robertson, 1977). Em populações de pequeno tamanho, espera-se que a seleção reduza a variabilidade nas linhas selecionadas. Porém, Falconer e Mackay (1996) notaram que, na prática, a variabilidade raramente parece declinar em experimentos de seleção. Existe, portanto, acréscimo na variabilidade genética, provavelmente causado pela mutação. Porém, mais pesquisas são necessárias para que se possa compreender as causas da manutenção da variabilidade em populações que estão sob seleção direcional.

Existe ainda uma dificuldade adicional relacionada ao fato das predições usuais estarem baseadas no modelo infinitesimal, embora estudos de mapeamento de QTL em *Drosophila*, camundongos e gado indiquem que existem regiões do genoma com efeitos substanciais na característica. Em experimentos de seleção em longo prazo com camundongos, Hill (2002) comenta que, de forma geral, o modelo infinitesimal tem se mostrado adequado em gerações iniciais de seleção. Porém, em gerações mais avançadas, verifica-se intensa redução na variância.

A resposta à seleção a longo prazo em populações finitas depende de diversos fatores, como deriva genética, desequilíbrio gamético de ligação, variância mutacional e tamanho efetivo da população. Todos esses fatores, por sua vez, dependem da característica em questão, da estrutura da população e da estratégia de seleção, que devem ser incorporados para se predizer a resposta à seleção. Ressalta-se também que a resposta em longo prazo é função da distribuição dos efeitos e das frequências gênicas de cada loco,

informação esta normalmente não disponível e que necessita ser considerada nas análises.

O número de locos num modelo de finitos locos deve ser grande o suficiente para garantir as propriedades do modelo infinitesimal, principalmente em relação à manutenção das frequências gênicas sob seleção. Boer e Van Arendonk (1992) trabalharam com modelo genético aditivo, considerando 64 ou 1.600 pares de locos bialélicos, não-ligados, visando estudar a predição dos efeitos aditivos em populações selecionadas e não selecionadas por cinco gerações. Ao realizar acasalamento ao acaso com o modelo de 64 pares de locos, os autores verificaram que: a variância genética aditiva foi reduzida devido ao estabelecimento de covariâncias entre os animais e ao aumento do coeficiente de endogamia; as predições dos efeitos genéticos aditivos foram empiricamente não-viesadas; a frequência média dos alelos favoráveis não se alterou na ausência de seleção. Porém, ao realizar seleção fenotípica, tanto para o modelo com 64, quanto para o modelo com 1.600 pares de locos, verificaram aumento na média dos valores genéticos aditivos simulados e decréscimo da variância genética aditiva, devido inclusive, às mudanças ocasionadas nas frequências gênicas. Como esperado, o aumento na frequência gênica dos alelos favoráveis foi menor na situação com 1.600 genes, sendo a variância genética aditiva mais próxima do valor esperado nas últimas gerações. Porém, notaram que para o modelo com 64 genes, as predições dos valores genéticos aditivos permaneceram não-viesadas em gerações mais avançadas quando a frequência inicial média dos alelos era 0,5, o que não ocorreu quando as frequências eram iguais a 0,2 ou 0,8. Esse fato é consequência da mudança na variância genética aditiva causada pela alteração da frequência gênica, que é ignorada na metodologia de modelos mistos, sendo relativamente maior com frequências gênicas extremas do que com frequências intermediárias (Falconer e Mackay, 1996).

Bulmer (1976) realizou um estudo sobre o efeito da seleção na variabilidade genética, utilizando dados simulados. A característica métrica estudada era determinada por 12 locos de efeitos aditivos, com dois alelos cada um, sendo que duas pressuposições acerca da ligação gênica foram realizadas: uma assumiu que todos os 12 locos localizavam-se em diferentes cromossomos e, portanto, segregavam independentemente; a outra, pressupôs

que havia quatro locos ligados em cada cromossomo, num total de três cromossomos. Esse autor considerou que a variância genética pode ser decomposta em três componentes: a variância genética propriamente dita, a variância resultante das covariâncias entre alelos do mesmo loco devido ao afastamento do Equilíbrio de Hardy-Weinberg e a variância resultante das covariâncias entre locos devido ao desequilíbrio de ligação. Sob acasalamento ao acaso em população finita, verificou que flutuação considerável pode ocorrer na variância genética, principalmente devido aos distanciamentos aleatórios do equilíbrio de ligação. Foi também verificado que existe uma correlação positiva entre os valores sucessivos do componente de variância devido ao desequilíbrio de ligação, indício de que as observações das variâncias genéticas em anos sucessivos não são independentes umas das outras.

O uso do modelo infinitesimal é questionado por diversos autores. Hill (1999) enfatiza que o modelo infinitesimal é apenas um modelo e não uma situação real, uma vez que o genoma não possui infinitos genes aditivos, não-ligados e de efeitos infinitesimalmente pequenos. Em diversas situações com gado, em que os dados obtidos são oriundos de poucas gerações, acrescenta que não é possível dizer quão bem esse modelo é apropriado. Fernando et al. (1994) apresentaram formulação alternativa do modelo misto (para característica influenciada por locos simples de grande efeito e por diversos locos com efeitos normalmente distribuídos, concomitantemente) com número finito de locos poligênicos, destacando que sua principal vantagem está no cálculo da verossimilhança, utilizando algoritmos eficientes desenvolvidos para modelos oligogênicos. Boer e Van Arendonk (1992) utilizaram modelo genético com 64 ou 1.600 pares de locos com dominância completa para estudar a predição dos efeitos aditivos e de dominância em populações selecionadas e não-selecionadas, ambas endogâmicas. Os autores verificaram que o uso do modelo estatístico que contém ambos os efeitos aditivos e de dominância, mas que ignora as mudanças na média e covariâncias genéticas associadas com a dominância devido à endogamia, resultou em predições significativamente viesadas de ambos os efeitos. Por outro lado, o modelo estatístico que considera o efeito médio da endogamia na média, enquanto ignora seus efeitos nas covariâncias genéticas associadas com a dominância, resultou em predições empiricamente não-viesadas.

Van Tassell et al. (1995) verificaram o efeito da seleção nas estimativas dos componentes de variância, utilizando a Amostragem de Gibbs num contexto Bayesiano (média e moda), a metodologia REML e o método MIVQUE em dados simulados. Todos os métodos, com exceção da moda da análise Bayesiana, geraram estimativas não viesadas, uma vez que toda a informação de parentesco foi incluída nas análises. Quando múltiplas gerações foram selecionadas, os autores verificaram que o quadrado médio do erro das variâncias genética aditiva e residual decresceu para as análises Bayesianas e para o REML ao ser comparado às situações de seleção aleatória e ausência de seleção. Resultados semelhantes foram encontrados por Pieramati e Van Vleck (1993). Porém, ambos os trabalhos utilizaram modelos bastante simplificados, onde não havia outro efeito fixo além da média geral. Schenkel (1998), por outro lado, verificou que num modelo em que foi incluído o efeito fixo de grupo contemporâneo, tanto a seleção fenotípica como a baseada no BLUP causaram um aumento no quadrado médio do erro das estimativas das variâncias genética e residual.

Experimentos de seleção artificial permitem testar o modelo infinitesimal, uma vez que é possível verificar se a resposta à seleção e as mudanças na variância genética estão de acordo com as predições feitas a partir da teoria genética. Experimentos de seleção com camundongos não têm apresentado resultados consistentes com relação à adequabilidade do modelo infinitesimal. Linhas selecionadas para consumo corrigido para peso corporal (Meyer e Hill, 1991) e particularmente, para aquelas selecionadas para peso magro em camundongos (Beniwal et al., 1992), mostraram fortes evidências de mudanças na variância genética aditiva, havendo falta de ajuste do modelo infinitesimal. Por outro lado, Martinez et al. (2000) estimaram satisfatoriamente a variância genética aditiva da população-base, mesmo após a população ter sido submetida a diversos ciclos de seleção para composição corporal, concluindo que não há, nesse caso, nenhum gene de efeito suficientemente grande capaz de produzir resultados diferentes dos esperados com relação às mudanças na média e na variância genética.

Thompson (2001) acrescenta que os desvios das predições dos valores genéticos sob o modelo infinitesimal indicam uma falha nesse modelo e podem auxiliar na compreensão dos mecanismos genéticos envolvidos na expressão

da característica e no desenvolvimento de modelos mais realísticos. Esse autor apresenta modificações que podem ser realizadas nos procedimentos de estimação de variâncias, capazes de incorporar as mudanças ocorridas na variância genética. Esses métodos foram utilizados em análises de populações selecionadas de camundongos (Beniwal et al., 1992; Heath et al., 1995), mostrando as mudanças na variância genética através das gerações. Segundo Hill (1999), a detecção de QTL de grande efeito também pode auxiliar na compreensão da falta de ajuste do modelo infinitesimal para determinadas características.

A maioria das teorias de seleção e análises para obtenção do BLUP, utilizadas no melhoramento animal, considera o modelo infinitesimal e atualmente extensões têm sido feitas para incluir um ou poucos QTL ligados a marcadores. Casos mais complexos de seleção também vêm sendo trabalhados, principalmente por Turelli e Barton (1994).

Modelos poligênicos finitos podem ser uma alternativa ao uso do modelo infinitesimal em avaliações genéticas com diversas gerações de seleção de características quantitativas. Bink (2002) propôs um método Bayesiano flexível que inclui o número de genes como uma variável aleatória no modelo poligênico finito e concluiu, com base nos resultados encontrados, que o modelo infinitesimal não deve ser substituído pelo poligênico finito, mas que ambos devem ser combinados com o número de genes tratado como aleatório, para que métodos mais robustos de estimação de componentes de variância em populações sob seleção possam ser desenvolvidos. Gonçalves (2003) testou o uso desses modelos combinados para estimar componentes de variância e o número de genes no modelo finito, utilizando metodologia Bayesiana. O autor evidenciou quatro genes de efeito principal, dois para gordura intramuscular e dois para espessura de toucinho em suínos.

2.1.1. Simulação Computacional de Sistemas Genéticos

A simulação consiste em construir um sistema que imite o funcionamento de uma realidade, com o objetivo de verificar o que aconteceria no sistema real, caso alterações de interesse fossem efetuadas no seu funcionamento

(Dachs, 1988). Informações importantes podem ser extraídas de um sistema simulado, com menor custo e maior rapidez.

A modelagem do processo seletivo, com suas complexas inter-relações, pode ser estudada por meio da simulação computacional. As simulações são baseadas num conjunto de pressuposições, das quais os dados são gerados e diferentes estruturas de seleção podem ser comparadas. As simulações computacionais de sistemas genéticos podem ser divididas em duas categorias: aquelas baseadas nos genes, ou seja, determina-se o tamanho de cromossomo, o número de locos, o número de alelos e assim por diante; e aquelas baseadas num infinito número de locos, nas quais médias e estruturas de variâncias são especificadas. Este último modelo é conhecido como modelo infinitesimal (Fisher, 1918) e o primeiro é chamado modelo de gene (Muir, 1997).

As simulações usuais sempre incluem um elemento estocástico e são geralmente referidas como simulações de Monte Carlo. A validade do modelo infinitesimal é claramente limitada, mas na seleção em curto prazo, caso a seleção não seja tão intensa e caso haja um grande número de locos aditivos influenciando a característica (maior que 500), as predições do modelo são razoáveis conforme verificado por simulação em nível de gene (Muir, 1997).

2.2. A Metodologia de Modelos Mistos

Usualmente, em problemas onde se deseja prever ou estimar parâmetros, assume-se que se tem disponível uma amostra (aleatória) de uma variável aleatória, cuja função densidade de probabilidade é supostamente conhecida, enquanto os parâmetros dessa distribuição são desconhecidos. O problema então é usar os valores observados da amostra para estimar ou prever os valores desconhecidos dos parâmetros.

Considere o seguinte modelo linear misto:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}, \quad [1]$$

em que:

\mathbf{y} = vetor de observações;

\mathbf{b} = vetor de efeitos fixos desconhecidos;

\mathbf{a} = vetor de efeitos aleatórios desconhecidos que representam os valores genéticos aditivos de cada animal;

\mathbf{e} = vetor de efeitos aleatórios residuais desconhecidos; e

\mathbf{X} e \mathbf{Z} = matrizes de incidência, que relacionam os registros aos efeitos fixos e aleatórios genéticos, respectivamente.

As pressuposições acerca da distribuição de \mathbf{y} , \mathbf{a} e \mathbf{e} podem ser descritas como:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{e} \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{ZGZ}' + \mathbf{R} & \mathbf{ZG} & \mathbf{R} \\ \mathbf{GZ}' & \mathbf{G} & \mathbf{F} \\ \mathbf{R} & \mathbf{F} & \mathbf{R} \end{bmatrix} \right\}, \quad [2]$$

em que

\mathbf{G} = matriz de variâncias e covariâncias dos efeitos aleatórios do vetor \mathbf{a} ;

\mathbf{R} = matriz de variâncias e covariâncias residuais.

As matrizes \mathbf{G} e \mathbf{R} são descritas como:

$$\bullet \quad \mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0, \quad [3]$$

em que:

\mathbf{A} = matriz que indica o grau de parentesco entre os indivíduos;

\mathbf{G}_0 = matriz de variâncias e covariâncias genéticas aditivas entre as características que compõem as observações;

\otimes = operador produto direto.

$$\bullet \quad \mathbf{R} = \mathbf{I} \otimes \mathbf{R}_0, \quad [4]$$

em que:

I = matriz identidade de ordem igual à dimensão linha de \mathbf{y} ;

R_0 = matriz de variâncias e covariâncias residuais entre as características que compõem as observações.

A Metodologia de Modelos Mistos (Henderson, 1973) vem sendo amplamente empregada no melhoramento animal, visando estimar efeitos fixos e prever valores genéticos em [1]. Na ausência de seleção e conhecendo-se as variâncias dos efeitos aleatórios do modelo, as Equações de Modelos Mistos de Henderson produzem o Melhor Estimador Linear Não-viesado (BLUE) das funções estimáveis de \mathbf{b} e o Melhor Preditor Linear Não-viesado (BLUP) dos efeitos aleatórios.

Os preditores lineares BLUP somente eram viáveis quando a ordem da matriz de variâncias e covariâncias de \mathbf{y} ($\mathbf{ZGZ}' + \mathbf{R} = \mathbf{V}$, conforme em [2]) era pequena o suficiente, de modo que pudesse ser invertida pelos programas computacionais disponíveis (Henderson, 1973). Portanto, a obtenção do BLUP envolvia grande esforço computacional, uma vez que, freqüentemente, \mathbf{V} era não-diagonal e de grandes dimensões.

Henderson et al. (1959) derivaram uma forma alternativa para obtenção do BLUP, de maneira semelhante à utilizada na derivação do método da Máxima Verossimilhança, pela maximização da função densidade de probabilidade conjunta das observações (\mathbf{y}) e dos valores genéticos (\mathbf{a}), em relação os efeitos fixos (\mathbf{b}) e aos valores genéticos (\mathbf{a}), obtendo assim as Equações de Modelos Mistos:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{\beta}^0 \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [5]$$

As soluções para os efeitos fixos podem ser obtidas pela absorção da segunda equação na primeira, de forma que:

$$\mathbf{\beta}^0 = \left\{ \mathbf{X}' \left[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1} \right] \mathbf{X} \right\}^{-1} \mathbf{X}' \left[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1} \right] \mathbf{y}. \quad [6]$$

Aplicando-se esse resultado à segunda equação obtém-se o BLUP de \mathbf{a} :

$$\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}^0). \quad [7]$$

O BLUP apresenta propriedades importantes, entre elas:

1) O BLUP maximiza a correlação entre as predições dos valores genéticos e os valores genéticos verdadeiros quando os animais são avaliados nas mesmas condições ambientais, uma vez que minimiza o erro de predição;

2) O BLUP é não-viesado, conforme requerido na sua derivação;

3) O BLUP de $\mathbf{L}'\mathbf{a}$, $\mathbf{L}'\hat{\mathbf{a}}$, é invariante para o valor de \mathbf{b} (\mathbf{L} é a matriz que orientou as decisões de seleção), ou seja, $\mathbf{L}'\hat{\mathbf{a}}$ é translação invariante. A invariância do BLUP de $\mathbf{L}'\mathbf{a}$ é consequência da restrição $\mathbf{b}\mathbf{X} = \mathbf{K}'$;

4) Na classe dos preditores lineares com média zero, o BLUP $\mathbf{L}'\hat{\mathbf{a}}$ maximiza a probabilidade de um correto ordenamento dos elementos de \mathbf{a} , o que confere ao método a qualidade de *melhor*.

Na ausência de seleção, caso as variâncias não sejam conhecidas, sendo necessário substituí-las por suas estimativas, estimadores e preditores não mais são os melhores. Porém, permanecem não-viesados se a distribuição dos dados for simétrica e se os estimadores dos componentes de variância forem de translação invariante. A seleção complica esse processo de diversas formas e suas consequências dependem do tipo de seleção que se opera.

2.2.1. Dados sob seleção

Entre as pressuposições para o uso das Equações de Modelos Mistos, como inicialmente proposto por Henderson (1973), está a condição de que os dados não sejam oriundos de seleção. Contudo, se a seleção ocorreu dentro dos níveis dos efeitos fixos, baseada na observação direta do fenótipo, ou se foi baseada no valor genético, mas conhece-se a matriz de parentesco de maneira completa, o BLUP continua com as mesmas propriedades.

Henderson (1975) derivou Equações de Modelos Mistos para obter a melhor predição linear não viesada dos valores genéticos para diferentes tipos de viés de seleção. Esses procedimentos pressupõem que:

- os vetores aleatórios do modelo seguem distribuição normal multivariada;
- matrizes de variâncias e covariâncias na população não selecionada são conhecidas;
- todos os dados utilizados para realizar as decisões de seleção são incluídos na análise;
- apenas um ciclo de seleção pode ter ocorrido. Goffinet (1983) provou, no entanto, que esses métodos de Henderson (1975) podem ser aplicados para vários ciclos de seleção.

Assim, é necessário construir a matriz que orientou as decisões de seleção, comumente designada matriz L . Caso essa matriz seja ortogonal à matriz de incidência dos efeitos fixos (X), as Equações de Modelos Mistos, como derivadas anteriormente, são suficientes para produzir o BLUE e o BLUP. Desta forma, modelos multivariados têm mostrado considerar o viés de seleção caso a seleção seja descrita como uma função de translação invariante das características que foram utilizadas nas decisões de seleção (Henderson, 1975). Da mesma forma, o viés ocasionado pelo acasalamento não aleatório dos animais pode ser contabilizado pela inclusão da matriz de parentesco completa até a população-base (Hudson e Schaeffer, 1983; Quaas e Pollak, 1980). A importância da matriz de parentesco, quando ela é completa e correta, em contabilizar os efeitos da seleção e do acasalamento não-aleatório nas avaliações genéticas vem sendo mostrada teoricamente e pela simulação (Pollak e Quaas, 1981; Sorensen e Kennedy, 1984b). A matriz de parentesco considera tanto o declínio na variância genética devido à deriva genética, caso todos os registros dos animais estejam disponíveis (Van der Werf e Boer, 1990), quanto a redução na variância causada pelo desequilíbrio gamético de ligação.

Sob o modelo infinitesimal (em particular, pressupondo que a seleção não produz mudanças na frequência gênica), a Metodologia de Modelo Misto considera totalmente os efeitos do desequilíbrio gamético, assim como da deriva genética (Sorensen e Kennedy 1983, 1984a). Isso ocorre, porque, sob o modelo infinitesimal, mesmo na presença de seleção e de deriva, a matriz de variâncias e covariâncias do vetor dos valores genéticos (G) permanece sendo

o produto da variância genética aditiva da população-base (σ_a^2) e da matriz dos numeradores dos coeficientes de parentesco (\mathbf{A}), ou seja, $\mathbf{G}=\mathbf{A}\sigma_a^2$, para análise com uma única característica. Contudo que a matriz de parentesco seja completa, esse resultado sugere que estimadores de componentes de variância que são suficientes podem produzir estimativas não-viesadas da variância genética aditiva da população-base. Essencialmente, uma matriz de parentesco que inclui os indivíduos não selecionados (geração zero) rastreia os valores genéticos dos indivíduos de qualquer geração de volta para a base.

Esse fato é consequência do comportamento dos resíduos da regressão (ε) do valor genético de um indivíduo (a_0) em relação aos valores genéticos de seu pai (a_s) e de sua mãe (a_m):

$$a_0 = \frac{1}{2}a_s + \frac{1}{2}a_m + \varepsilon \quad [8]$$

Esse resíduo, também conhecido como amostragem Mendeliana, resulta da segregação dos alelos nos locos heterozigotos dos pais. Sob o modelo infinitesimal, ele independe dos valores genéticos dos pais e possui média zero e variância $\frac{1}{2}\sigma_a^2(1-\bar{F})$, em que \bar{F} é o coeficiente de endogamia médio do pai e da mãe (Thompson, 1977). Porém, se esse modelo genético não se aplica, os valores residuais podem realmente variar conforme os valores genéticos dos pais, tendo a seleção influência sobre a distribuição desses resíduos. Porém, se a mudança na frequência gênica for de pequena magnitude ao longo do processo seletivo, o viés pode não ser tão expressivo.

Dada a sensibilidade da Metodologia de Modelos Mistos ao cumprimento de suas pressuposições (em particular, o modelo infinitesimal), alguma forma de validação do modelo é necessária para aplicar esta metodologia com confiança. Uma forma é verificar se as estimativas da variância genética permanecem estáveis conforme são acrescentadas gerações de seleção ao processo (Meyer e Hill, 1991). Caso o modelo infinitesimal seja adequado, a matriz de parentesco contabiliza as mudanças na variância genética nessas gerações de seleção adicionadas. Por outro lado, se a variância genética está

sendo alterada de uma forma não explicável pelo modelo infinitesimal, utilizando-se dados de gerações adicionais de seleção, pode-se obter estimativas bastante diferentes da variância genética aditiva da população-base. Da mesma forma, se a mesma população-base é usada para formar linhas de controle e de seleção, espera-se obter a mesma estimativa da variância genética da população-base a partir de ambas as linhas. A não ocorrência dessas condições indica falha do modelo em questão (Meyer e Hill, 1991).

2.3. Método da Máxima Verossimilhança Restrita

A predição dos valores genéticos exige o conhecimento das magnitudes dos componentes de variância genética e residual da característica em questão. Como, freqüentemente, essas variâncias não são conhecidas, o recurso é estimá-las primeiramente e depois, predizer os valores genéticos de interesse, a partir do mesmo conjunto de dados. Nessa circunstância, os resultados obtidos utilizando-se as Equações de Modelos Mistos (Henderson et al., 1959) não são BLUP nem BLUE, ao menos que, por acaso, as estimativas usadas forem exatamente os valores reais. Conforme essas estimativas se aproximam das reais variâncias, os resultados preditos se aproximam do BLUP e do BLUE. A questão então é como melhor estimar esses componentes de variâncias.

Atualmente, o método mais utilizado em programas de melhoramento genético para estimar componentes de variância é o REML. Esse método de estimação é uma modificação do método de Máxima Verossimilhança (ML), o qual não leva em consideração a perda de graus de liberdade causada pela estimação dos efeitos fixos do modelo. Esse fato pode ser preocupante quando o modelo inclui efeitos fixos com muitos níveis. O método REML foi então proposto por Patterson e Thompson (1971), baseando-se na projeção dos dados num subespaço livre dos efeitos fixos, onde é feita a maximização da verossimilhança. Como a obtenção das estimativas de **G** e **R** envolve cálculo de derivadas da verossimilhança, dificultando o processo de estimação, Smith e Graser (1986) e Graser et al. (1987) propuseram o método livre de derivadas

(DF), no qual as estimativas dos componentes de variância são obtidas sem que haja necessidade de se calcular derivadas e esperanças, sendo obtidas por tentativas diversas de valores de \mathbf{G} e \mathbf{R} , até que a combinação que maximiza o logaritmo da verossimilhança seja encontrado (Boldman et al., 1995). A equação a ser maximizada é a seguinte:

$$\Lambda = -0,5[(n-p) \log(2\pi) + \log|\mathbf{R}| + \log|\mathbf{G}| + \log|\mathbf{C}| + \mathbf{y}' \mathbf{P} \mathbf{y}], \quad [9]$$

em que:

n = número de registros;

p = posto da matriz de coeficientes relativo aos efeitos fixos;

\mathbf{C} = matriz de coeficientes de posto completo das Equações de Modelos Mistos; e

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \quad [10]$$

Na expressão [9], a constante é ignorada e a cada rodada os outros quatro termos precisam ser calculados.

Do ponto de vista prático, o estimador de máxima verossimilhança é uma importante ferramenta para o pesquisador. A máxima verossimilhança permite obter estimadores com propriedades desejáveis (assintoticamente). Trata-se de um estimador de translação invariante (função de toda estatística suficiente), consistente e assintoticamente normal e eficiente; e produz estimativas não-negativas (porém, neste caso, torna-se viesado). Por outro lado, requer uma solução iterativa, não havendo garantia de convergência, podendo, ainda, atingir um máximo local e não global.

2.3.1. Dados sob seleção

Assim como o BLUP e BLUE, as estimativas REML das variâncias podem ser viesadas pela seleção. Em particular, se a população-base consiste de indivíduos não selecionados e não correlacionados, então, sob o modelo infinitesimal, o método REML produz essencialmente estimativas não-viesadas da variância genética aditiva na população-base (Henderson et al., 1959;

Sorensen e Kennedy, 1984b; Gianola e Fernando, 1986; Gianola et al., 1988). Sorensen e Kennedy (1986) utilizaram o método MIVQUE (Estimador Quadrático Não-Viesado de Variância Mínima) para estimar componentes de variância em dados simulados sob seleção e verificaram que o método gerou estimativas não viesadas da variância genética aditiva da população-base, quando se utilizou um modelo animal e a matriz de parentesco completa foi incluída nas análises. Juga e Thompson (1989) verificaram, por meio de dados simulados, que o viés nas estimativas da herdabilidade pode ser considerável após algumas gerações de seleção, caso não seja utilizado método apropriado. Os resultados encontrados por esses autores sugerem que a inclusão da matriz completa de parentesco ao utilizar o método REML permite a obtenção de estimativas não viesadas dos componentes de variância.

Os trabalhos de simulação acima mencionados consideraram como adequado o modelo infinitesimal, e, portanto, a aproximação da normalidade multivariada ainda pôde ser considerada. Porém, com finito número de locos, particularmente com genes de grande efeito, a pressuposição de normalidade é menos sustentável (Sorensen e Kennedy, 1984a). Na verdade, o resíduo da regressão em [8] não é mais independente dos valores genéticos dos pais e a expressão $\mathbf{G} = \mathbf{A}\sigma_a^2$ não mais se aplica. Logo, os estimadores baseados nos modelos mistos serão viesados. A relevância dos problemas associados com número de locos finitos depende do tipo de modelo genético considerado. Serão acentuados com genes de grande efeito com frequências extremas ou com locos apresentando dominância multidirecional. Segundo Sorensen e Kennedy (1984a), estudos de simulação com vários modelos de finitos locos podem produzir resultados bastante interessantes nessa área.

Diversos autores (Van der Werf e de Boer, 1990; Van der Werf, 1992) mostraram que, quando a população-base consiste de indivíduos previamente selecionados, o método REML pode fornecer estimativas viesadas da variância genética aditiva, mesmo se todo o pedigree dos indivíduos até a população-base for incluído. Assim, uma importante questão levantada é como proceder quando a população-base está, ela mesma, sob seleção. Graser et al. (1987) sugeriram tratar os valores genéticos dos indivíduos da população-base como efeitos fixos, ao invés de aleatórios. A justificativa para essa sugestão é que se

os pais são oriundos de seleção, eles não são uma amostra aleatória da população-base. Uma vez que as estimativas REML não são viesadas para efeitos fixos, qualquer viés na variância da amostra inicial seria ignorado ao considerar os valores genéticos dos pais dessa população como fixos. Estudos de simulação, no entanto, mostraram que mesmo o viés sendo reduzido nessa situação, a seleção nas gerações seguintes introduz viés adicional (Van der Werf, 1992). Breda (2003) avaliou o comportamento da variância genética aditiva em populações selecionadas, utilizando dados simulados e obteve estimativas viesadas das variâncias genéticas aditivas, assim como da eficiência de seleção baseada no BLUP em relação à seleção fenotípica. A autora também concluiu que quanto mais distante da população-base verdadeira, maior a tendência de superestimar o valor da variância genética aditiva e da eficiência de seleção.

Mesmo sob o modelo infinitesimal, a variância genética aditiva muda ao longo do tempo. Enquanto o método REML fornece uma estimativa da variância genética aditiva da população-base (que é não viesada, caso as pressuposições do modelo se apliquem), ele não provê imediatamente uma estimativa da variância genética em alguma particular geração de seleção. A metodologia mais direta é usar a regressão pais-filhos para cada geração de seleção para estimar a variância genética aditiva na geração dos pais. Segundo Walsh (2000), a desvantagem dessa metodologia está relacionada ao pequeno tamanho amostral tipicamente associado a cada geração, resultando em grandes desvios-padrão para cada estimativa da variância ou da herdabilidade. Seria, portanto, mais adequado combinar as informações de todas as gerações para melhorar as estimativas desejadas. Sorensen e Kennedy (1984a) sugeriram que uma forma que combina informações para estimar a variância numa determinada geração t é utilizar a Metodologia de Modelos Mistos, considerando essa geração em questão como sendo a população-base. Em particular, consideram-se somente os dados da geração t em diante e a matriz de parentesco é ajustada, de maneira a pressupor que a geração t é a população-base. Porém, estudos de simulação feitos por Van der Werf e de Boer (1990) mostraram que essa metodologia tende a superestimar a real variância. Outra forma seria usar a variância entre os valores genéticos preditos dentro de cada geração. Apesar de parecer uma sugestão razoável,

existem complicações inerentes ao método: adiciona-se uma incerteza nos valores genéticos preditos; a variância genética usada para obter as predições BLUP sofre forte influência das próprias predições realizadas, havendo uma circularidade e, portanto, essa metodologia não é recomendada (Walsh, 2000). Atualmente, não existe uma solução satisfatória para o problema de combinar informações ao longo das gerações para estimar a variância numa geração em particular.

2.4. Análise Bayesiana

Recentemente, métodos Bayesianos têm sido propostos como alternativas de análise genética no melhoramento animal. Apesar da teoria Bayesiana estar fundamentada nos trabalhos desenvolvidos pelo reverendo Thomas Bayes em 1761, somente nas últimas décadas surgiram publicações mostrando a sua aplicação em trabalhos na área de melhoramento genético. Gianola e Foulley (1982) introduziram métodos Bayesianos na análise de características de limiar e, posteriormente, Gianola e Fernando (1986) destacaram possibilidades adicionais do uso das técnicas Bayesianas no melhoramento animal. Ronningen (1971) e Dempfle (1977) já haviam apresentado, há alguns anos, que o BLUP poderia ser interpretado como um preditor Bayesiano, e Harville (1974) forneceu uma interpretação Bayesiana do REML.

No melhoramento animal, como comentado anteriormente, duas etapas são realizadas na avaliação genética: primeiramente, estimam-se os componentes de (co)variância e depois obtém-se o BLUE e o BLUP dos efeitos fixos e aleatórios, respectivamente, sendo os valores paramétricos das variâncias substituídos por suas estimativas usualmente baseadas na máxima verossimilhança. Sob seleção ao acaso, esse procedimento deve convergir em probabilidade para o BLUE e o BLUP conforme se aumenta a informação sobre os componentes de variância na amostra; porém, suas propriedades freqüentistas sob seleção não-aleatória são desconhecidas. Uma deficiência desse procedimento é que os erros de estimação dos componentes de dispersão não são levados em consideração na predição dos valores genéticos (Sorensen et al., 1994; Wang et al., 1994a; Wang et al., 1994b).

Conforme a teoria da probabilidade associada aos métodos Bayesianos, as inferências devem ser baseadas nas distribuições marginais *a posteriori* dos parâmetros de interesse, de forma que a incerteza sobre os demais parâmetros desconhecidos seja totalmente levada em consideração. Deve-se, inicialmente, especificar a distribuição conjunta *a posteriori* de todos os parâmetros desconhecidos. A partir dessa distribuição conjunta, a distribuição marginal *a posteriori* do parâmetro de interesse, como por exemplo, do valor genético de um animal, é obtida pela integração sucessiva em relação a todos os parâmetros de perturbação, sendo estes os efeitos fixos, todos os demais efeitos aleatórios excetuando-se aquele de interesse, além dos componentes de (co)variância. No entanto, a resolução dessas múltiplas integrais normalmente é de alta complexidade ou mesmo analiticamente inviável. Nesse contexto, os métodos da Cadeia de Markov Monte Carlo (MCMC) vêm sendo aplicados para se estimar distribuições marginais *a posteriori*.

Os métodos MCMC são uma família de processos iterativos para geração aproximada de amostras de distribuições multivariadas. Nesses métodos, o conhecimento da variável aleatória gerada só é relevante para a geração da variável seguinte, sendo portanto, irrelevante na predição das variáveis futuras. Esses métodos se tornaram uma importante ferramenta computacional na estatística Bayesiana, uma vez que geram uma cadeia de Markov por meio da simulação iterativa de Monte Carlo, cuja distribuição estacionária é a própria densidade *a posteriori* de interesse (Sorensen e Gianola, 2002). Um método MCMC bastante utilizado é o Algoritmo de Metropolis-Hastings (Hastings, 1970), sendo a Amostragem de Gibbs (Geman e Geman, 1984) um caso especial desse algoritmo que vem sendo empregado por diversos autores na avaliação genética animal e vegetal, inclusive na análise de dados longitudinais, de características de limiar e de locos de características quantitativas.

Segundo Gianola e Foulley (1990), o método REML apresenta desvantagens, uma vez que as estimativas obtidas são os elementos do vetor modal da distribuição *a posteriori* conjunta dos componentes de variância. Do ponto de vista teórico, a regra de decisão ótima de Bayes sob perda quadrática é a média *a posteriori*, e não, a moda *a posteriori*. Além disso, se as inferências

sobre um único componente de variância são desejadas, a distribuição marginal desse componente deve ser usada ao invés da distribuição conjunta de todos os componentes. Dentro desse contexto, esses autores propuseram um método que, fornecendo as distribuições *a priori* e a verossimilhança que gerou os dados, constrói a distribuição conjunta *a posteriori* de todos os parâmetros. A distribuição marginal de um único componente de variância é obtida pela integração da função conjunta em relação a todos os outros parâmetros contidos no modelo. Assim, média, moda, mediana e variância podem ser obtidas. Probabilidades sobre o parâmetro podem ser calculadas e intervalos de confiança Bayesianos podem ser construídos, fornecendo solução completa para o problema de estimação de componentes de variância. Como, na prática, a integração da distribuição muitas vezes não pode ser realizada analiticamente, deve-se recorrer aos métodos numéricos.

Nas análises REML, a estimativa pontual da herdabilidade é considerada implicitamente para se obter as respostas à seleção. Gianola et al. (1986) mostraram que uma alternativa para se estimar a resposta à seleção seria o uso da distribuição marginal da herdabilidade. Sorensen e Johansson (1992) acrescentam que essa estratégia parece ser mais relevante quando a amostra é pequena e, portanto, não contém informação suficiente, o que é bastante comum no caso de experimentos de seleção.

Ambas as escolas, freqüentista e Bayesiana, utilizam a verossimilhança na realização de suas inferências. De um lado, a verossimilhança assume um papel central na inferência Bayesiana como a função que expressa toda a informação proveniente dos dados. Pelo outro lado, o método da Máxima Verossimilhança possui interessantes propriedades freqüentistas, como mencionado anteriormente. Assim, quando o conjunto de dados é suficientemente grande, os resultados são bastante similares na maioria dos casos (Blasco, 2001).

2.4.1. Teorema de Bayes

Considere um vetor não observável θ (θ é o vetor dos parâmetros que se deseja estimar), um vetor de observações \mathbf{y} e a função densidade conjunta $f(\theta, \mathbf{y})$. Conforme a teoria da probabilidade:

$$f(\theta, \mathbf{y}) = f(\mathbf{y}|\theta) f(\theta) \quad [11]$$

e

$$f(\theta, \mathbf{y}) = f(\theta|\mathbf{y}) f(\mathbf{y}), \quad [12]$$

onde $f(\theta)$ e $f(\mathbf{y})$ são as densidades marginais de θ e \mathbf{y} , respectivamente. Conforme [11] e [12], tem-se que:

$$f(\theta|\mathbf{y}) = f(\mathbf{y}|\theta) f(\theta) / f(\mathbf{y}) \quad [13]$$

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

$$f(\mathbf{y}) = \int_{R_\theta} f(\mathbf{y}, \theta) d\theta = \int_{R_\theta} f(\mathbf{y}|\theta) f(\theta) d\theta = E_\theta [f(\mathbf{y}|\theta)], \quad [14]$$

onde E_θ representa a média com respeito à distribuição de θ . Nota-se, portanto, que $f(\mathbf{y})$ não é uma função de θ . Logo, [13] pode ser escrita em forma de proporcionalidade (Gianola e Fernando, 1986):

$$f(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta) f(\theta) \quad [15]$$

De acordo com a terminologia Bayesiana, $f(\theta)$ é conhecida como a densidade *a priori* de θ , que reflete a incerteza relativa sobre os possíveis valores de θ antes das observações serem realizadas, ou seja, antes de se conhecer os valores do vetor \mathbf{y} . A densidade $f(\mathbf{y}|\theta)$ é a função de verossimilhança, que representa a contribuição de \mathbf{y} para o conhecimento de

θ . Finalmente, $f(\theta|\mathbf{y})$ é a densidade *a posteriori* que considera o estado de incerteza sobre q dado um conhecimento prévio. Inferências sobre θ são feitas a partir da densidade *a posteriori* e a solução para o problema estatístico é finalizada com:

$$\text{Prob}(\theta \in R|\mathbf{y}) = \int_R f(\theta|\mathbf{y}) d\theta, \quad [16]$$

onde R é a região do espaço de θ . Fixando a probabilidade na expressão [16] em, por exemplo, $1 - \alpha$, para um dado α , é possível obter um intervalo para θ , de forma que seu conteúdo de probabilidade seja $1 - \alpha$.

A metodologia Bayesiana não se baseia no conhecimento abstrato da amostragem repetida dos dados conforme requerem os procedimentos clássicos. Nestes, as propriedades dos estimadores pontuais e de intervalos são baseados nos dados que se deveria obter caso o experimento fosse repetido um número infinito de vezes, e não nos dados que se têm disponíveis (Blasco, 2001). Na análise Bayesiana, as inferências são feitas diretamente a partir da distribuição *a posteriori* [15], ou seja, somente os dados observados e a densidade *a priori* são necessários. Quando diferentes conclusões são obtidas ao se utilizar *priors* alternativos, os resultados da análise devem ser vistos com cautela, porque há um indício de que a própria função de verossimilhança é pouco informativa, podendo sofrer influência de qualquer *prior*. Wang et al. (1994a) comentam que essa é uma das vantagens da análise Bayesiana, pois evita que conclusões equivocadas sejam feitas baseadas em verossimilhanças fracas.

O teorema de Bayes fornece uma solução precisa para o problema de tamanhos de amostras finitos (Gianola e Fernando, 1986). Para cada conjunto de dados, que pode ser grande ou pequeno, existe uma distribuição *a posteriori* exata da qual pode-se realizar as inferências. Muitos procedimentos estatísticos amplamente utilizados no melhoramento animal, como a estimação pelo método da máxima verossimilhança, somente são justificados assintoticamente, ou seja, suas propriedades são bem definidas quando as amostras são grandes. Nessas situações, a densidade *a posteriori* tende para a função de verossimilhança, havendo uma justificativa Bayesiana para

procedimentos de máxima verossimilhança. Porém, esse não é o caso quando o tamanho amostral é finito, situação na qual somente a expressão [15] fornece uma solução conceitual para o problema (Gianola e Fernando, 1986).

2.4.2. Seleção e Acasalamento Não Aleatório

Os dados de campo, sobre os quais a avaliação genética é realizada, usualmente provêm de rebanhos onde a seleção é rotineiramente praticada. A questão de como estimar valores genéticos em populações submetidas à seleção é de grande importância prática.

No contexto Bayesiano pode-se mostrar que, em determinadas condições, as inferências sobre os valores e parâmetros genéticos devem ser feitas a partir da densidade *a posteriori* construída como se a seleção não tivesse ocorrido. O resultado é concordante com os desenvolvimentos baseados na normalidade e aplica-se a qualquer distribuição, conforme mostram Gianola e Fernando (1986).

Considere \mathbf{y}_0 sendo um vetor de méritos desconhecidos e \mathbf{z}_2 e \mathbf{z}_3 , os vetores de parâmetros secundários desconhecidos, ou seja, de *efeitos fixos* e das (co)variâncias, respectivamente. A seleção e as decisões de acasalamento são feitas com base nos dados disponíveis \mathbf{y}_0 , um vetor $n_0 \times 1$, que pode conter informações de uma ou várias características. Se \mathbf{y}_0 pertence a i -ésima partição (R_i) discreta e disjunta do espaço n_0 -dimensional, seleção e decisões de acasalamento são feitas, as quais levam à realização de \mathbf{y}_i ($i = 1, 2, \dots, s$). Então $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_s$ são conjuntos de dados mutuamente exclusivos porque $\text{Prob}(\mathbf{y}_0 \in R_i \text{ e } \mathbf{y}_0 \in R_j) = 0, i \neq j$.

Na ausência de seleção e de acasalamento não aleatório, é útil considerar que $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_s, \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$ possuem distribuição conjunta com densidade:

$$f(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_s, \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3). \quad [17]$$

A densidade marginal de $f(\mathbf{y}_0, \mathbf{y}_i, \theta_1, \theta_2, \theta_3)$ é obtida sob apropriada integração. As inferências devem ser baseadas em:

$$f(\theta_1, \theta_2, \theta_3 | \mathbf{y}_0, \mathbf{y}_i), \quad [18]$$

e a função pode ser integrada em relação a θ_2 e θ_3 :

$$f(\theta_1 | \mathbf{y}) = \int_{R_{\theta_2}} \int_{R_{\theta_3}} f(\theta_1 | \theta_2, \theta_3, \mathbf{y}) \times f(\theta_2, \theta_3 | \mathbf{y}) d\theta_2 d\theta_3. \quad [19]$$

Sendo $\hat{\theta}_3$ o estimador modal, as inferências são feitas a partir de:

$$f(\theta_1 | \mathbf{y}) \doteq f\left(\theta_1 \mid \theta_3 = \hat{\theta}_3, \mathbf{y}\right) \quad [20]$$

Agora, quando a seleção e as decisões de acasalamentos são realizadas, a densidade *a posteriori* de $\mathbf{y}_0, \mathbf{y}_i, \theta_1, \theta_2, \theta_3$ é definida como se $\mathbf{y}_0 \in R_i$. Então, escreve-se:

$$f(\mathbf{y}_0, \mathbf{y}_i, \theta_1, \theta_2, \theta_3 | \mathbf{y}_0 \in R_i) = \frac{f(\mathbf{y}_0, \mathbf{y}_i, \theta_1, \theta_2, \theta_3)}{\text{Prob}(\mathbf{y}_0 \in R_i)} \quad [21]$$

porque $\mathbf{y}_0 \in R_i$ é um evento discreto. Utilizando argumentos de probabilidade condicional:

$$f(\mathbf{y}_0, \mathbf{y}_i | \mathbf{y}_0 \in R_i) = \frac{f(\mathbf{y}_0, \mathbf{y}_i)}{\text{Prob}(\mathbf{y}_0 \in R_i)} \quad [22]$$

onde $f(\mathbf{y}_0, \mathbf{y}_i)$ é a densidade conjunta de $\mathbf{y}_0, \mathbf{y}_i$ como se a seleção não tivesse ocorrido. Então, combinando [21] e [22]:

$$f(\alpha_1, \alpha_2, \alpha_3 | \mathbf{y}_0, \mathbf{y}_i, \mathbf{y}_0 \in R_i) = \frac{f(\mathbf{y}_0, \mathbf{y}_i, \alpha_1, \alpha_2, \alpha_3) / \text{Prob}(\mathbf{y}_0 \in R_i)}{f(\mathbf{y}_0, \mathbf{y}_i) / \text{Prob}(\mathbf{y}_0 \in R_i)} = f(\alpha_1, \alpha_2, \alpha_3 | \mathbf{y}_0, \mathbf{y}_i), \quad [23]$$

a qual é precisamente a expressão [18]. Isso prova que se a seleção e as decisões de acasalamento são baseadas em \mathbf{y}_0 e, ainda, se \mathbf{y}_0 está presente nos dados disponíveis para análise, então as inferências sobre os valores genéticos, efeitos fixos e variâncias e covariâncias devem ser feitas a partir da densidade *a posteriori*, como se a seleção não tivesse ocorrido. Esse fato generaliza os resultados de Henderson (1975), os quais dependem da normalidade. Caso os parâmetros secundários α_3 sejam desconhecidos e caso o interesse esteja em α_1 , então utiliza-se o argumento em [20]: empregando-se todos os dados $(\mathbf{y}_0, \mathbf{y}_i)$, estima-se α_3 com um estimador modal (estimador de máxima verossimilhança restrita caso pressuponha-se normalidade e utilize-se *flat priors*) e faz-se as inferências a partir de:

$$f\left(\alpha_1 | \mathbf{y}_0, \mathbf{y}_i, \alpha_3 = \hat{\alpha}_3\right). \quad [24]$$

Nota-se que no contexto Bayesiano, o requerimento da translação invariante, feita por Henderson (1975) não é necessário, uma vez que na metodologia Bayesiana α_2 é considerado como variável aleatória. A não ser por essa diferença, [24] está em completa concordância com Henderson (1975) quando se pressupõe normalidade.

Concluindo, a distribuição *a posteriori* dos parâmetros e valores genéticos é a mesma com ou sem seleção ou acasalamento não aleatório. Logo, fornecidos as observações e o conhecimento *a priori*, qualquer regra de decisão baseada na distribuição *a posteriori*, incluindo \mathbf{y}_0 , não será afetada pelo acasalamento não aleatório dos animais.

2.4.3. Métodos Bayesianos Aplicados

Apesar da inferência do ponto de vista Bayesiano estar bem estruturada, a sua aplicação era, até recentemente, bastante problemática. A metodologia

Bayesiana necessita da resolução de integrais multidimensionais complicadas ou o uso de aproximações mais ou menos acuradas. O uso recente das técnicas MCMC vem resolvendo a maioria desses problemas (Blasco, 2001), apesar de também ter gerado outros, relacionados com a convergência das cadeias. As técnicas MCMC fornecem amostras aleatórias das distribuições *a posteriori* conjuntas ou marginais e então, a média, o desvio-padrão, os intervalos de confiança podem ser estimados diretamente dessas amostras sem que seja necessária a resolução de integrais.

2.4.3.1. Distribuição *a posteriori* completa

As inferências Bayesianas são realizadas com base nas densidades marginais *a posteriori* dos parâmetros de interesse. Essas, por sua vez, são obtidas a partir da distribuição conjunta *a posteriori* completa, em que todos os efeitos considerados no modelo devem ser incluídos. A obtenção dessa distribuição conjunta no contexto do melhoramento animal será mostrada a seguir.

Considere o modelo linear misto [1]. O objetivo é obter amostras aleatórias das distribuições marginais *a posteriori* de todos os parâmetros desconhecidos, ou seja, de:

$$f(\beta_1|\mathbf{y}), f(\beta_2|\mathbf{y}), \dots, f(a_1|\mathbf{y}), f(a_2|\mathbf{y}), \dots, f(s_a^2|\mathbf{y}), f(s_e^2|\mathbf{y}) \quad [25]$$

e também para qualquer combinação desses parâmetros, como por exemplo:

$$f[s_a^2/(s_a^2 + s_e^2)|\mathbf{y}] = f(h^2|\mathbf{y}). \quad [26]$$

Essas amostras serão utilizadas para inferências. Para obtê-las, primeiramente, acessam-se amostras aleatórias da distribuição conjunta *a posteriori* completa:

$$f(\boldsymbol{\beta}, \mathbf{a}, s_a^2, s_e^2 | \mathbf{y}). \quad [27]$$

Essa função conduzirá a uma matriz na qual cada linha será uma amostra de cada parâmetro desconhecido, e cada coluna, um conjunto de pontos amostrados aleatoriamente da distribuição marginal *a posteriori* de cada parâmetro desconhecido.

Para se obter [27], a função de verossimilhança e as distribuições *a priori* são necessárias conforme visto em [15], uma vez que:

$$f(\boldsymbol{\beta}, \mathbf{a}, s_a^2, s_e^2 | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\beta}, \mathbf{a}, s_a^2, s_e^2) f(\boldsymbol{\beta}, \mathbf{a}, s_a^2, s_e^2). \quad [28]$$

Assume-se normalmente que as distribuições *a priori* dos parâmetros de dispersão são mutuamente independentes entre eles e independentes das distribuições *a priori* dos valores genéticos e de ambiente. Pressupõe-se também que as distribuições *a priori* para b e \mathbf{a} são independentes. Assim:

$$f(\boldsymbol{\beta}, \mathbf{a}, s_a^2, s_e^2 | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\beta}, \mathbf{a}, s_a^2, s_e^2) f(\boldsymbol{\beta}) f(\mathbf{a}) f(s_a^2) f(s_e^2). \quad [29]$$

Pressupõe-se que \mathbf{y} segue distribuição normal multivariada:

$$\mathbf{y} | \boldsymbol{\beta}, \mathbf{a}, s_a^2, s_e^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a}, \mathbf{I} s_e^2). \quad [30]$$

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

$$f(\boldsymbol{\beta}) \propto \text{constante}. \quad [31]$$

Para o vetor dos valores genéticos aditivos, assume-se que seus elementos seguem distribuição normal multivariada:

$$\mathbf{a}|\mathbf{A}, \mathbf{s}_a^2 \sim N(0, \mathbf{A}\mathbf{s}_a^2). \quad [32]$$

Para completar a descrição do modelo na análise Bayesiana, as distribuições *a priori* dos componentes de variância precisam ser especificadas. Assume-se, normalmente, que ambos os componentes do modelo seguem uma distribuição de qui-quadrado invertida escalonada, de forma que:

$$p(s_i^2 | \nu_i, s_i^2) \propto (s_i^2)^{-\left(\frac{\nu_i+1}{2}\right)} \exp\left(-\frac{\nu_i s_i^2}{2s_i^2}\right) \quad (i = a, e), \quad [33]$$

em que ν_i e s_i^2 são os hiperparâmetros da distribuição.

Assim, a forma da distribuição conjunta *a posteriori* [29] é representada por:

$$f(\boldsymbol{\beta}, \mathbf{a}, \mathbf{s}_a^2, \mathbf{s}_e^2 | \mathbf{y}, \mathbf{n}_i, \mathbf{s}_i^2) \propto (\mathbf{s}_e^2)^{-\left(\frac{n+n_e+1}{2}\right)} \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{a})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{a}) + \nu_e \mathbf{s}_e^2}{2\mathbf{s}_e^2}\right] \\ \times (\mathbf{s}_a^2)^{-\left(\frac{q+n_a+1}{2}\right)} \exp\left[-\frac{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + n_a \mathbf{s}_a^2}{2\mathbf{s}_a^2}\right], \quad [34]$$

onde n e q se referem ao número total de observações e o número de elementos do vetor \mathbf{a} , respectivamente.

2.4.3.2. Amostragem de Gibbs

Normalmente, não é possível obter amostras aleatórias diretamente da distribuição *a posteriori* completa [34] porque se trata de uma função de alta

dimensão, sendo o produto de diversas funções multidimensionais. Para reduzir a dimensionalidade do problema e tornar possível obter as amostras aleatórias, técnicas MCMC são usadas. A mais empregada tem sido a Amostragem de Gibbs, que é uma técnica voltada para geração de vetores aleatórios de uma distribuição conjunta por meio da amostragem sucessiva das distribuições condicionais de todas as variáveis aleatórias presentes no modelo.

Para implementação da Amostragem de Gibbs, deve-se derivar todas as distribuições condicionais *a posteriori* a partir da conjunta apresentada em [34], cujo resultado pode ser encontrado em Wang et al. (1994a) e Sorensen et al. (1994). As distribuições condicionais completas de cada parâmetro do modelo são chamadas Amostradores de Gibbs.

O procedimento consiste em retirar amostras das distribuições condicionais (Amostradores de Gibbs) de uma forma iterativa, como será ilustrado a seguir.

Escreve-se a função condicional:

$$f(\beta_1 | \beta_2, \beta_3, \dots, \beta_i, a_1, a_2, \dots, a_i, s_a^2, s_e^2, \mathbf{y}), \quad [35]$$

fornecendo, em seguida, valores arbitrários para $\beta_2, \beta_3, \dots, \beta_i, a_1, a_2, \dots, a_i, s_a^2, s_e^2$. Amostra-se, então, um valor aleatório para β_1 dessa distribuição condicional. Para isto, a distribuição condicional deve ser de um tipo para os quais algoritmos de amostragem estão disponíveis; caso contrário, outras técnicas MCMC devem ser utilizadas, como o algoritmo de Metropolis-Hastings. Agora, com o valor de β_1 aleatoriamente amostrado e os demais valores arbitrários, tira-se uma amostra aleatória de β_2 da distribuição condicional:

$$f(\beta_2 | \beta_1, \beta_3, \dots, \beta_i, a_1, a_2, \dots, a_i, s_a^2, s_e^2, \mathbf{y}). \quad [36]$$

Com os valores amostrados para β_1 e β_2 e os demais valores arbitrários iniciais, retira-se uma amostra aleatória de β_3 , e procede-se dessa forma até que uma amostra de cada parâmetro desconhecido tenha sido tomada de cada distribuição condicional. Após esse processo, inicia-se outro, repetindo-se todo o ciclo, utilizando como valores iniciais as amostras aleatórias obtidas para cada parâmetro desconhecido. Após alguns ciclos de iteração, as amostras aleatórias extraídas das condicionais (conjuntas ou marginais) são também amostras aleatórias da distribuição (conjunta ou marginal) *a posteriori* (Geman e Geman, 1984). Maiores detalhes sobre a Amostragem de Gibbs podem ser encontrados em Casella e George (1992).

Van Tassell et al. (1995) destacaram vantagens do uso da Amostragem de Gibbs em relação aos outros métodos de estimação de componentes de variância, citados a seguir: não é necessária a obtenção das soluções de modelos mistos; quando técnicas simples de matriz esparsa são usadas, arquivos de dados maiores do que os utilizados pelo REML são possíveis; são geradas estimativas diretas e exatas dos componentes de variância e valores genéticos e intervalos de confiança para essas estimativas; trata-se de método apropriado para ser utilizado em microcomputadores e estações de trabalho, devido à necessidade de armazenar quantidade relativamente pequena de informação na memória.

3. MATERIAL E MÉTODOS

3.1. Simulação dos dados

Os dados utilizados na elaboração desse trabalho foram simulados utilizando-se o pacote computacional GENESYS (Euclides, 1996), composto por nove programas, escritos em linguagem FORTRAN. O Genesys se baseia na simulação em nível de gene, permitindo a elaboração de genomas de certa complexidade para estudos de métodos de seleção, testes de pressuposições e avaliações de novas metodologias de seleção. Maiores detalhes sobre o pacote são apresentados por Euclides e Guimarães (1997).

Quatro diferentes tipos de populações foram simulados:

1^a) **GRAN 900**: População grande, sendo cada geração constituída por 2.400 indivíduos com registros e 1.000 pais, em que a característica sob seleção é governada por 900 pares de locos;

2^a) **GRAN 10**: População grande, sendo cada geração constituída por 2.400 indivíduos com registros e 1.000 pais, em que a característica sob seleção é governada por 10 pares de locos;

3^a) **PEQ 900**: População pequena, sendo cada geração constituída por 120 indivíduos com registros e 50 pais, em que a característica sob seleção é governada por 900 pares de locos;

4^a) **PEQ 10**: População pequena, sendo cada geração constituída por 120 indivíduos com registros e 50 pais, em que a característica sob seleção é governada por 10 pares de locos.

Inicialmente, foram simulados quatro distintos genomas para formação de cada uma das populações acima mencionadas. Para os genomas simulados foram estabelecidos locos bialélicos, cujas freqüências iniciais foram simuladas com base na distribuição normal, com valor médio inicial igual a 0,50.

Quadro 1 - Parâmetros utilizados no processo de simulação das quatro populações

Descrição	GRAN 900	GRAN 10	PEQ 900	PEQ 10
Tamanho do genoma (<i>cM</i>)	3.000	1.000	3.000	1.000
Nº de cromossomos	20	10	20	10
Nº de pares de locos	900	10	900	10
Nº de efeitos fixos	2	2	2	2
Nº de indivíduos na pop. básica	3.000	3.000	3.000	3.000
População-Base (G₀):				
Nº de machos	200	200	10	10
Nº de fêmeas/macho	4	4	4	4
Nº de filhos/fêmea/macho	3	3	3	3
Média Genética	185,9	183,6	187,4	179,0
Variância Genética	146,0	1.009,2	132,2	882,7
Herdabilidade	0,32	0,35	0,31	0,32

Apenas os efeitos aditivos dos locos quantitativos foram simulados conforme distribuição normal, sendo ignorados os efeitos de dominância e epistasia. As frequências gênicas iniciais foram iguais em ambos os sexos. Foram simulados dois efeitos fixos, *A* e *B*, possuindo o primeiro 15 e o segundo seis subclasses. Os efeitos de ambiente não-controláveis foram simulados com base na distribuição normal. O processo de simulação empregado permitiu a geração de ligações gênicas entre determinados locos nas populações em que a característica é governada por 900 genes e, contrariamente, gerou locos não ligados nas populações em que a característica é governada por apenas 10 genes. Características adicionais sobre os genomas encontram-se no Quadro 1.

Posteriormente, com base nos parâmetros descritos no Quadro 1, foram formadas quatro populações básicas com 3.000 indivíduos cada, compostas por 1.500 machos e 1.500 fêmeas. A partir dessas populações básicas, os indivíduos foram selecionados ao acaso, originando quatro diferentes Populações-Base (G₀). Assim, as Populações-Base GRAN 900 e GRAN 10 foram formadas pela seleção ao acaso de 200 machos e 800 fêmeas, sendo produzidos três descendentes por fêmea, originando populações com 2.400 indivíduos; as Populações-Base PEQ 900 e PEQ 10 foram formadas pela seleção ao acaso de 10 machos e 40 fêmeas, sendo produzidos três descendentes por fêmea, originando populações com 120 indivíduos.

A partir da População-Base, dois processos seletivos foram aplicados: Seleção ao Acaso (SAA), onde os indivíduos foram selecionados aleatoriamente e Seleção Fenotípica (SF), onde os indivíduos foram selecionados com base no seu valor fenotípico (considerou-se que a seleção visa aumentar a média da população). Foram avançadas 10 gerações discretas para cada um dos processos seletivos, em que o número de machos e fêmeas selecionados, assim como o de descendentes deixados a cada geração foram iguais aos apresentados na formação da População-Base (Quadro 1). O esquema de todo o processo de simulação acima descrito encontra-se na Figura 1. Para as populações GRAN 900 e GRAN 10 foram feitas 300 repetições e para as populações PEQ 900 e PEQ 10 foram realizadas 500 repetições para cada um dos processos seletivos na obtenção das 10 gerações. Assim, os valores genéticos e ambientais reais para cada indivíduo a cada geração são valores médios obtidos de 300 ou 500 repetições, conforme o tamanho da população.

3.2. Análises dos Valores Reais

3.2.1. Curtose, Assimetria e Teste de Normalidade

Os efeitos genético-aditivos e ambientais reais foram submetidos a análises que sugerem o quanto suas respectivas distribuições se afastam da normalidade. Os valores de curtose, assimetria, assim como o teste de Normalidade de Kolmogorov-Smirnov foram obtidos para as populações das gerações G_0 e G_{10} , que sofreram Seleção ao Acaso e Seleção Fenotípica, utilizando-se o pacote estatístico SAS[®] (Littell, 1991). Valores de curtose e assimetria próximos de zero indicam distribuições mais próximas da mesocúrtica e mais simétricas, respectivamente. Testes de normalidade significativos sugerem que a hipótese dos dados seguirem distribuição normal deve ser rejeitada.

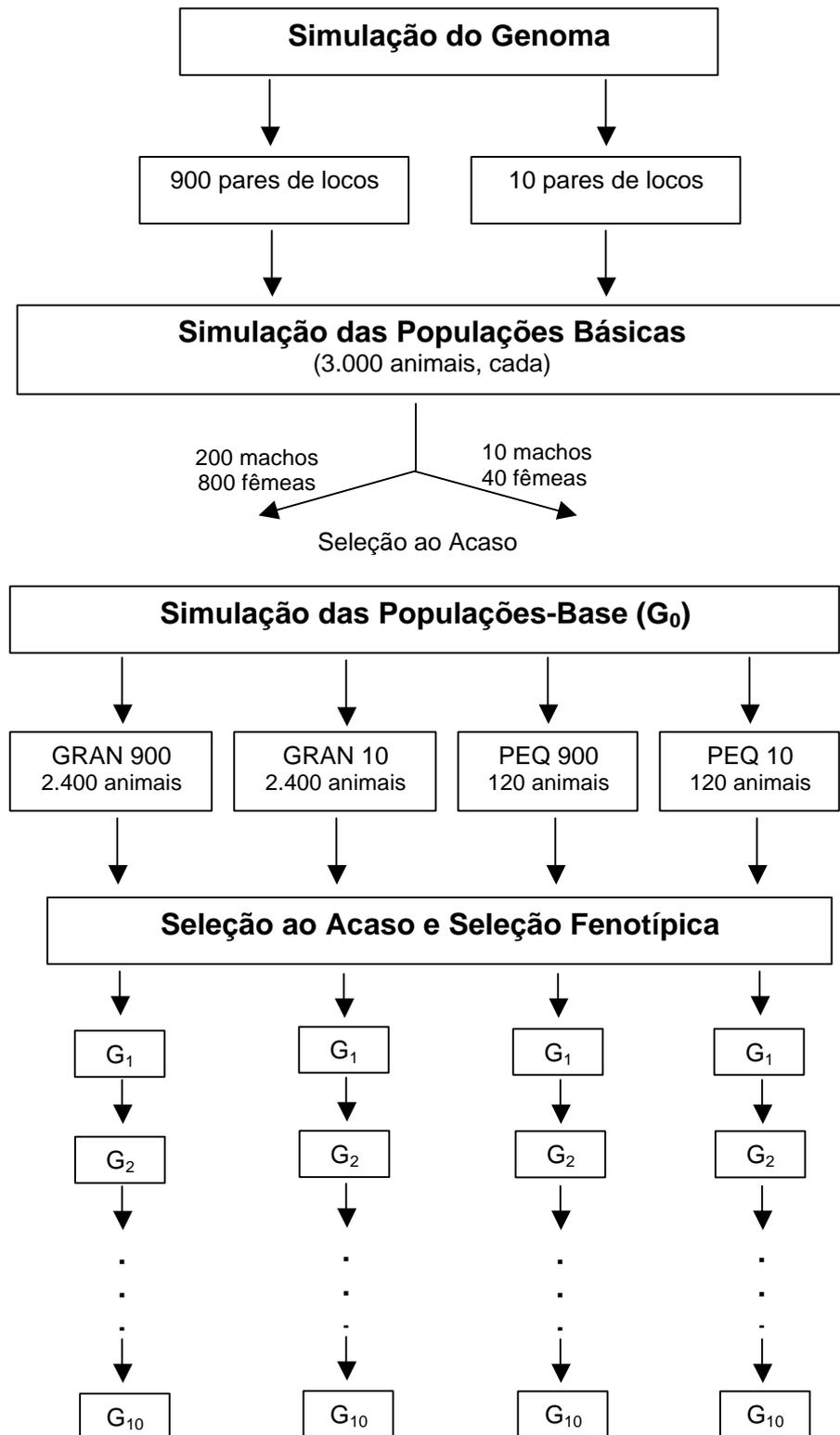


Figura 1 - Esquema empregado na simulação de quatro tipos de populações a serem analisadas: GRAN 900, GRAN 10, PEQ 900 e PEQ 10.

3.2.2. Componentes de Variância

Os valores reais dos componentes de variância genética aditiva e ambiental ao longo das 10 gerações de Seleção ao Acaso e de Seleção Fenotípica foram submetidos à análise visual, possibilitando verificar e comparar o comportamento de tais componentes nas diferentes situações estudadas. O cálculo das variâncias reais (baseado nos valores médios de 300 ou 500 repetições) a cada geração foi realizado com base na seguinte expressão:

$$\hat{s}_i^2 = \frac{\sum_{j=1}^n (w_i)_j^2 - \left[\frac{\sum_{j=1}^n (w_i)_j}{n} \right]^2}{n-1} \quad (i=a, e), \quad [37]$$

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

3.3. Avaliação Genética

Duas metodologias foram empregadas na avaliação genética dos dados simulados, visando estimar os componentes de variância e prever os valores genéticos dos indivíduos: *BLUP Empírico* ou *EBLUP*, em que a Metodologia dos Modelos Mistos foi aplicada, utilizando-se as estimativas dos componentes de variância obtidas pela Máxima Verossimilhança Restrita (REML) e *Análise Bayesiana*, utilizando a Amostragem de Gibbs.

3.3.1. Estruturas Consideradas

Os componentes de variância foram estimados por ambas as metodologias para as seguintes situações:

1) \mathbf{G}_0 : foram utilizados os registros dos indivíduos da Geração 0 para estimar os componentes de variância da Geração 0;

2) **G₁₀ (SAA)**: foram utilizados os registros dos indivíduos da Geração 10 de Seleção ao Acaso e informação de parentesco sem registro dos pais dos indivíduos da Geração 10, para estimar os componentes de variância da Geração 0;

3) **G₁₀ (SF)**: foram utilizados os registros dos indivíduos da Geração 10 de Seleção Fenotípica e informação de parentesco sem registro dos pais dos indivíduos da Geração 10, para estimar os componentes de variância da Geração 0;

4) **G₀₋₁₀ (SAA)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 0 de Seleção ao Acaso, para estimar os componentes de variância da Geração 0;

5) **G₀₋₁₀ (SF)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 0 de Seleção Fenotípica, para estimar os componentes de variância da Geração 0.

Na predição dos valores genéticos, as seguintes estruturas foram consideradas:

1) **G₀**: foram utilizados os registros dos indivíduos da Geração 0 para predizer os valores genéticos dos indivíduos da Geração 0;

2) **G₁₀ SAA**: foram utilizados os registros dos indivíduos da Geração 10 de Seleção ao Acaso e informação de parentesco sem registro dos pais dos indivíduos da Geração 10, para predizer os valores genéticos dos indivíduos da Geração 10;

3) **G₁₀ SF**: foram utilizados os registros dos indivíduos da Geração 10 de Seleção Fenotípica e informação de parentesco sem registro dos pais dos indivíduos da Geração 10, para predizer os valores genéticos dos indivíduos da Geração 10;

4) **G₁₀ SAA (0-10)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 0 de Seleção ao Acaso, para predizer os valores genéticos dos indivíduos da Geração 10;

5) **G₁₀ SF (0-10)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 0 de Seleção Fenotípica, para predizer os valores genéticos dos indivíduos da Geração 10;

6) **G₁₀ SF (4-10)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 4 de Seleção Fenotípica, para prever os valores genéticos dos indivíduos da Geração 10;

7) **G₁₀ SF (7-10)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 7 de Seleção Fenotípica, para prever os valores genéticos dos indivíduos da Geração 10;

8) **G₁₀ SF (8-10)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 8 de Seleção Fenotípica, para prever os valores genéticos dos indivíduos da Geração 10;

9) **G₁₀ SF (9-10)**: foram utilizados todos os registros e informação de parentesco completa da Geração 10 até a Geração 9 de Seleção Fenotípica, para prever os valores genéticos dos indivíduos da Geração 10.

3.3.2. Análise Bayesiana *via* Amostragem de Gibbs

A análise Bayesiana foi realizada utilizando-se o programa MTGSAM – *Multiple Trait Gibbs Sampling in Animal Models* (Van Tassell e Van Vleck, 1995). O MTGSAM é composto por um conjunto de programas escritos em FORTRAN, desenvolvido para estimação de componentes de variância por meio da Amostragem de Gibbs.

3.3.2.1. Modelo

O seguinte modelo linear misto unicaracterístico foi utilizado na análise dos dados simulados:

$$y_{ijk} = \mu + A_i + B_j + a_{ijk} + e_{ijk}, \quad [38]$$

em que y_{ijk} é a observação fenotípica do $k^{\text{ésimo}}$ indivíduo no $i^{\text{ésimo}}$ efeito *fixo* A e no $j^{\text{ésimo}}$ efeito *fixo* B; μ é uma constante inerente a todas as observações; A_i é o efeito do $i^{\text{ésimo}}$ efeito *fixo* A; B_j é o efeito do $j^{\text{ésimo}}$ efeito *fixo* B; a_{ijk} é o efeito aleatório genético aditivo do $k^{\text{ésimo}}$ indivíduo no $i^{\text{ésimo}}$ efeito *fixo* A e no $j^{\text{ésimo}}$ efeito *fixo* B; e_{ijk} é o efeito aleatório ambiental. Matricialmente, este modelo

pode ser representado como em [1], pressupondo-se que $\mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2)$ e $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, conforme descrito anteriormente.

3.3.2.2. Distribuições *a priori*

Na análise Bayesiana, torna-se necessário especificar também as distribuições *a priori* para todos os efeitos do modelo. O programa MTGSAM foi desenvolvido utilizando *flat priors* (distribuições uniformes) para os efeitos *fixos*, ou seja, considerando-se que há um vago conhecimento *a priori* sobre esses efeitos, como apresentado em [31]. Pressupõe que os efeitos aleatórios, inclusive os ambientais, seguem distribuição normal, sendo que para os efeitos genéticos existe uma pressuposição adicional sobre uma conhecida estrutura de covariâncias, representada pela matriz de parentesco. Adicionalmente, pressupõe que os componentes de variância genética e ambiental seguem distribuição de qui-quadrado invertida escalonada (Sorensen et. al., 1994) :

$$p(\mathbf{s}_i^2 | \mathbf{n}_i, s_i^2) \propto (\mathbf{s}_i^2)^{-\left(\frac{\mathbf{n}_i}{2} + 1\right)} \exp\left(-\frac{\mathbf{n}_i \mathbf{s}_i^2}{2s_i^2}\right) \quad (i = a, e), \quad [39]$$

em que

\mathbf{s}_i^2 = componente de variância i ;

\mathbf{n}_i = parâmetro da distribuição que corresponde ao “grau de confiança” do componente de variância i ;

s_i^2 = parâmetro da distribuição que pode ser interpretado como um valor *a priori* do componente de variância i .

A derivação das distribuições condicionais completas, necessárias para a implementação da Amostragem de Gibbs, é encontrada em Van Tassell e Van Vleck (1995).

Três níveis de informação *a priori* em relação aos componentes de variância foram considerados nessa metodologia: não-informativo (NI), pouco informativo (PI) e informativo (I). Os valores dos hiperparâmetros da distribuição encontram-se no Quadro 2. Para o nível não-informativo esses valores resultam em distribuição uniforme para as variâncias.

Quadro 2 - Valores dos hiperparâmetros da distribuição de qui-quadrado invertida escalonada dos componentes de variância (CV) genética (s_a^2) e ambiental (s_e^2), conforme o nível de informação considerado: não-informativo (NI), pouco informativo (PI) e informativo (I)

População	CV	Hiperparâmetro	NI	PI	I
GRAN 900	s_a^2	n_a	-2	4	90
		s_a^2	0	146,04	146,04
	s_e^2	n_e	-2	8	100
		s_e^2	0	304,16	304,16
GRAN 10	s_a^2	n_a	-2	6	90
		s_a^2	0	1.009,20	1.009,20
	s_e^2	n_e	-2	12	140
		s_e^2	0	1.915,20	1.915,20
PEQ 900	s_a^2	n_a	-2	4	90
		s_a^2	0	132,23	132,23
	s_e^2	n_e	-2	6	100
		s_e^2	0	283,51	283,51
PEQ 10	s_a^2	n_a	-2	6	120
		s_a^2	0	882,67	882,67
	s_e^2	n_e	-2	10	160
		s_e^2	0	1.853,58	1.853,58

Para os níveis pouco informativo e informativo, os valores do hiperparâmetro s_i^2 correspondem aos valores reais dos componentes de variância e os valores do hiperparâmetro n_i foram definidos fixando-se o valor de s_i^2 desejado e variando-se os de n_i , de forma que maior ou menor grau de confiança fosse atribuído às variâncias informadas. Este procedimento foi realizado por meio da análise gráfica das distribuições de qui-quadrado invertida escalonada.

3.3.2.3. Cadeia de Gibbs

Uma única cadeia de Gibbs foi gerada para cada análise. A inspeção visual das amostras geradas a cada iteração, assim como os programas

GIBBSIT (Raftery e Lewis, 1992) e *GIBANAL* (Van Kaam, 1998), foram utilizados para definir o tamanho da cadeia, o período de *burn-in* e o intervalo de amostragem (*thinning interval*) para obtenção das densidades marginais a *posteriori*. Para as populações GRAN 900 e GRAN 10, o tamanho da cadeia variou de 35.000 a 70.000 amostras e o período de *burn-in* de 5.000 a 10.000 iterações. Para a população PEQ 900, o tamanho da cadeia variou de 35.000 a 10.000.000 amostras e o período de *burn-in* de 5.000 a 5.000.000 iterações e para a população PEQ 10, o tamanho da cadeia variou de 120.000 a 10.000.000 amostras e o período de *burn-in* de 20.000 a 5.000.000 iterações.

3.3.3. EBLUP

O mesmo modelo utilizado na Análise Bayesiana foi considerado na estimação dos componentes de variância pelo REML e predição dos valores genéticos empregando-se a Metodologia de Modelos Mistos. Foram utilizados para este fim os programas do pacote MTDFREML – *Multiple Trait Derivative-Free Restricted Maximum Likelihood* (Boldman et. al., 1995).

3.3.4. Critérios de Comparação das Metodologias

3.3.4.1. Componentes de Variância

Porcentagem de Erro

A Porcentagem do Erro (*PE*) obtida na estimação dos componentes de variância genética aditiva e ambiental foi calculada para as populações das gerações G_0 , G_{10} e G_0 a G_{10} , tanto para Seleção ao Acaso quanto para Seleção Fenotípica. Como referência, foram utilizados os valores reais dos respectivos componentes da População-Base (G_0). Assim:

$$PE = \frac{(\hat{\sigma}_i^2 - \sigma_i^2)100}{\sigma_i^2} \quad (i=a,e), \quad [40]$$

em que $\hat{\sigma}_i^2$ é o componente de variância estimado e σ_i^2 é o componente de variância real da População-Base.

3.3.4.2. Valores Genéticos

Com base nos valores genéticos preditos pela metodologia EBLUP e pela metodologia Bayesiana com *priors* não-informativo e informativo, foram calculados o Quadrado Médio do Desvio, a Porcentagem de Seleccionados em Comum e a Correlação de Ordem, visando comparar as metodologias empregadas, assim como as diferentes estruturas consideradas.

Quadrado Médio do Desvio

O Quadrado Médio do Desvio (*QMD*) dos valores genéticos preditos em relação aos reais foram calculados como a seguir:

$$QMD = \sum_{i=1}^n \frac{1}{n} (\hat{a}_i - a_i)^2, \quad [41]$$

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

Porcentagem de Seleccionados em Comum

A porcentagem de indivíduos em comum, seleccionados entre os 15% melhores indivíduos, considerando o ordenamento real e o obtido pelos valores genéticos preditos, foi calculada para as estruturas de 1 a 5, conforme descrito no item **3.3.1**, para os valores genéticos.

Correlação de Ordem

A Correlação de Ordem ou de Spearman entre os valores genéticos reais e preditos foi calculada, utilizando-se o pacote estatístico SAS[®] (Littell, 1991).

3.3.5. Adequabilidade do Modelo Infinitesimal

A adequabilidade do modelo infinitesimal como pressuposição nas análises genéticas foi verificada conforme o critério adotado por Meyer e Hill (1991), em que todas as estimativas da variância genética aditiva provenientes das análises, considerando o acréscimo sucessivo de gerações, devem ser idênticas, caso o modelo infinitesimal seja adequado. Assim, as variâncias obtidas a cada análise devem ser a da população-base. Foram utilizadas nesse estudo as estimativas obtidas pelo método REML.

3.3.6. Análise genética em gerações avançadas

A Porcentagem de Erro das estimativas dos componentes de variância genética aditiva, assim como o Quadrado Médio do Desvio e a Correlação de Ordem entre valores genéticos reais e preditos, obtidos pelo EBLUP e utilizando somente dados de gerações mais avançadas, foram calculados para as situações em que ocorreu Seleção Fenotípica. Assim, diferentes populações-base foram consideradas como referência nas análises: populações da Geração 4, da Geração 7, da Geração 8 e da Geração 9. Para cada uma das situações, foram incluídos todos os registros, assim como a matriz de parentesco da Geração 10 até a *nova* referência da população-base, conforme apresentado no item 3.3.1.

4. RESULTADOS E DISCUSSÃO

4.1. Análises dos Valores Reais

4.1.1. Curtose, Assimetria e Teste de Normalidade

Os valores de curtose e assimetria e os resultados do teste de normalidade para as Gerações 0 e 10 quando se realizou Seleção ao Acaso e Seleção Fenotípica encontram-se no Quadro 3.

Verifica-se que, para a população GRAN 900, os valores de curtose e assimetria para os efeitos genético aditivo e ambiental permaneceram bem próximos de zero, com exceção da curtose para o efeito ambiental que apresentou valor levemente superior em relação aos demais para a G_0 (0,17) e levemente inferior (-0,15) para a G_{10} com Seleção Fenotípica. O teste de normalidade para esta população foi não significativo para ambos os efeitos e para as gerações analisadas. Esses resultados indicam que, mesmo após 10 gerações de seleção (ao acaso e fenotípica), os valores genéticos e ambientais continuam seguindo distribuição normal.

Para população GRAN 10, no entanto, os resultados encontrados indicam que os valores genéticos não seguem distribuição normal, considerando as gerações G_0 e G_{10} de Seleção ao Acaso e de Seleção Fenotípica. A curtose e a assimetria do efeito genético aditivo apresentaram maiores problemas quando comparados com GRAN 900, chegando, respectivamente, aos valores 0,26 e -0,85 na Geração 10 para Seleção Fenotípica. Adicionalmente, os testes de normalidade para este efeito foram todos significativos para as gerações analisadas. Nota-se que, mesmo para uma população com 2.400 indivíduos, caso a característica seja governada por poucos locos com efeitos mais expressivos, os valores genéticos da própria população-base já não seguem

Quadro 3 - Curtose, assimetria e teste de normalidade dos efeitos genético (a) e ambiental (e) reais para as populações referentes às Gerações 0 e 10 para Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

População	Curtose		Assimetria		Teste de Normalidade	
	a	e	a	e	a	e
GRAN 900						
G ₀	0,01	0,17	0,01	0,02	ns	ns
G ₁₀ SAA	0,02	-0,05	-0,04	0,00	ns	ns
G ₁₀ SF	0,03	-0,15	-0,03	-0,04	ns	ns
GRAN 10						
G ₀	-0,17	-0,15	-0,03	-0,04	**	ns
G ₁₀ SAA	-0,04	-0,09	-0,18	0,00	**	ns
G ₁₀ SF	0,26	0,05	-0,85	0,08	**	ns
PEQ 900						
G ₀	0,10	0,21	-0,02	-0,23	ns	ns
G ₁₀ SAA	-0,30	-0,39	-0,25	-0,05	ns	ns
G ₁₀ SF	0,05	-0,34	-0,16	-0,15	ns	ns
PEQ 10						
G ₀	0,19	-0,09	0,33	0,08	ns	ns
G ₁₀ SAA	0,49	-0,43	0,28	0,47	**	*
G ₁₀ SF	-0,35	0,16	-0,57	-0,27	**	*

* - significativo a 5% de probabilidade pelo teste de Kolmogorov-Smirnov.

** - significativo a 1% de probabilidade pelo teste de Kolmogorov-Smirnov.

ns – não-significativo a 5% de probabilidade pelo teste de Kolmogorov-Smirnov.

distribuição normal, permanecendo esse problema para as gerações mais avançadas de seleção. Para os efeitos ambientais, os valores de curtose e assimetria foram, em geral, mais próximos de zero e os testes de normalidade foram não-significativos para todas as gerações, indicando que, para este efeito, o número de locos que governa a característica, assim como a Seleção Fenotípica por 10 gerações não alteram sua distribuição, pressupostamente normal.

Verifica-se na população PEQ 900 que maiores valores absolutos de curtose e assimetria foram encontrados em relação à população GRAN 900, tanto para os efeitos genético-aditivos quanto para os efeitos ambientais. No entanto, os testes de normalidade foram todos não significativos, indicando que, o elevado número de locos com pequenos efeitos está sendo suficiente para não alterar significativamente a forma da distribuição.

Para a população PEQ 10, apesar do teste de normalidade ser não-significativo para os efeitos genético-aditivos na G_0 , nota-se que, nessa situação, os valores de curtose e assimetria foram os de maior valor absoluto quando comparados com as demais populações. Na G_{10} de Seleção ao Acaso e de Seleção Fenotípica, a curtose e a assimetria foram bastante expressivas, sendo os testes de normalidade significativos, tanto para os efeitos genético-aditivos quanto para os efeitos ambientais. Conforme os resultados obtidos, nota-se que o pequeno número de locos com efeitos expressivos associado com pequeno número de indivíduos (120) ocasiona desvios da normalidade das distribuições dos efeitos genéticos e ambientais em gerações mais avançadas de seleção.

A análise visual dos desvios da normalidade dos efeitos genético-aditivos por meio de histogramas (Figura 2) confirma os resultados encontrados no Quadro 3. Percebe-se ajuste adequado dos efeitos à normalidade para GRAN 900, o que não ocorre para GRAN 10, sendo mais acentuado o problema na situação em que há Seleção Fenotípica. Verifica-se, para população PEQ 900, que os efeitos não apresentaram ajustes tão bons quanto foram para GRAN 900, o que se reflete nos valores de curtose e assimetria apresentados no Quadro 3. Em relação à população PEQ 10, observa-se maiores problemas para as populações que sofreram seleção, principalmente na situação em que ocorreu Seleção Fenotípica. Esses resultados confirmam que o número de locos que governa a característica influencia no desvio da normalidade, principalmente dos valores genéticos aditivos, tanto na população-base quanto em gerações mais avançadas, provenientes de seleção.

4.1.2. Componentes de Variâncias Reais

Os valores reais dos componentes de variância genética aditiva e ambiental ao longo das 10 gerações (além da população-base) encontram-se nas Figuras 3 e 4, respectivamente. Em relação à variância genética aditiva, para as situações em que ocorreu Seleção ao Acaso, nota-se que não há tendência de queda, ao comparar G_0 com G_{10} (na população GRAN 900 passou de 146,0 para 161,2 e na PEQ 900 passou de 132,2 para 194,8).

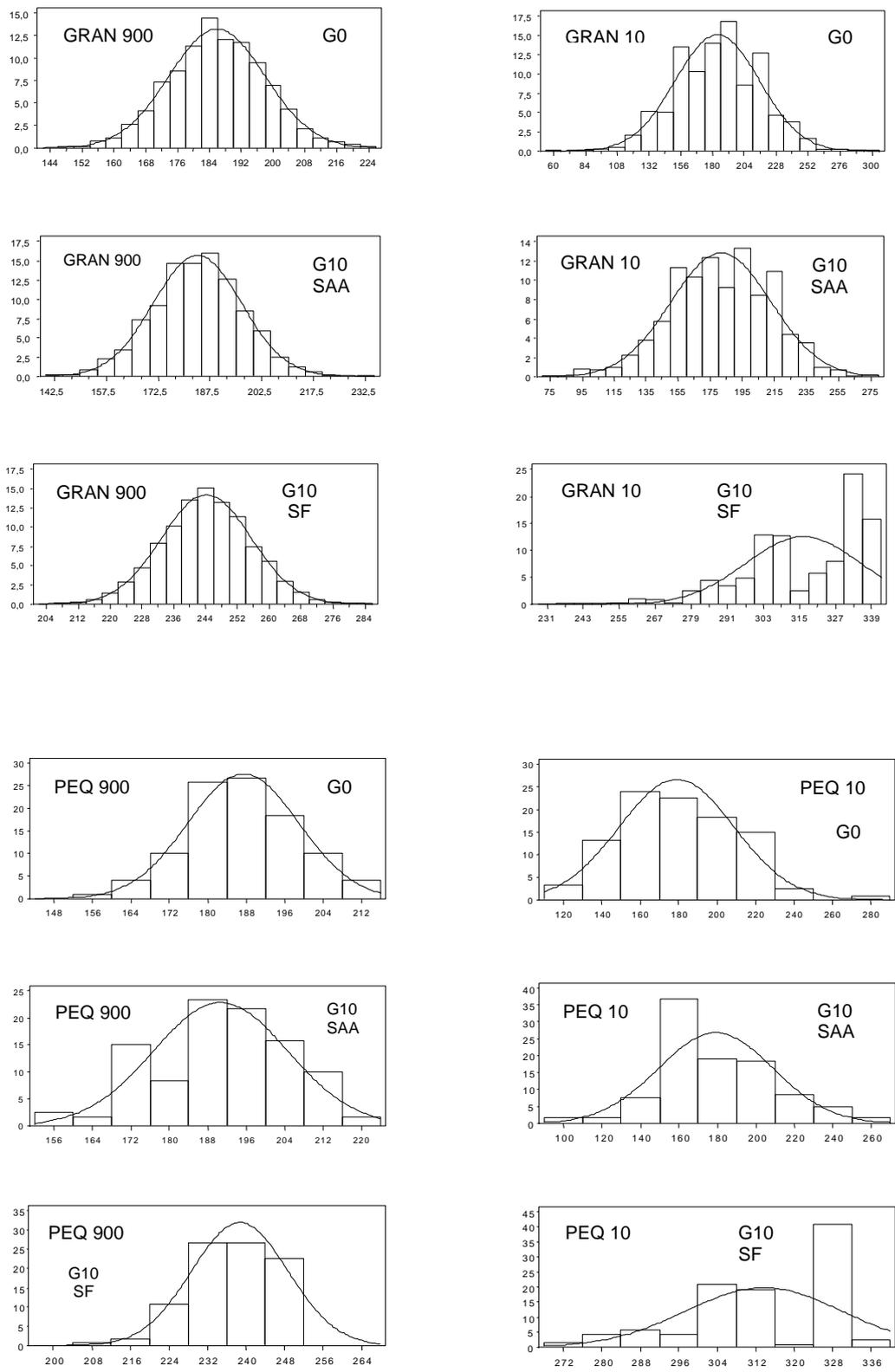


Figura 2 - Distribuição dos valores genéticos das populações GRAN 900, GRAN 10, PEQ 900 e PEQ 10 das gerações G₀ e G₁₀, para Seleção ao Acaso (SAA) e Seleção Fenotípica (SF).

Nas situações em que ocorreu Seleção Fenotípica, a variância genética aditiva decresceu em todas as populações. Porém, percebe-se claramente a queda mais acentuada para as populações GRAN 10 e PEQ 10. A análise de dados sob seleção usualmente se baseia no modelo infinitesimal (Fisher, 1918), que pressupõe que a característica é governada por infinitos genes de efeito aditivo, não-ligados, e que esses genes têm influência infinitesimalmente pequena na expressão da característica. Dessa forma, mesmo sob seleção capaz de alterar a média genética, mudanças na variância aditiva devido a mudanças na frequência gênica são consideradas desprezíveis. Nesse contexto, a diminuição da variância aditiva seria causada pela endogamia (em populações de tamanho finito a variância dentro de família é reduzida conforme aumenta o coeficiente de endogamia) e pelo desequilíbrio gamético (correlação negativa entre pares de locos) e, caso o processo seletivo fosse interrompido, essa variância seria recuperada nas gerações seguintes (Bulmer, 1971). Conforme os resultados encontrados, a queda acentuada na variância aditiva nas populações com 10 pares de locos está sendo causada não somente pelo aumento da endogamia e pelo desequilíbrio gamético, mas também pela influência das mudanças nas frequências gênicas. Com um número finito de locos, a seleção também leva a mudanças na frequência gênica, ocasionando alterações permanentes na variância genética.

Vale ressaltar que as populações com 900 pares de locos possuem ainda uma fonte adicional de variação pelo fato de possuírem genes ligados, que provavelmente causam um aumento no grau de desequilíbrio gamético devido à seleção (Martinez et al., 2000). Por outro lado, os 10 pares de locos das outras duas populações não estão ligados. Apesar dessa diferença, o pequeno número de locos mostra-se responsável pela mudança mais acentuada nas variâncias aditivas dessas últimas ao serem comparadas com as populações com 900 pares de locos. Esses resultados confirmam a idéia de diversos autores (Meyer e Hill, 1991; Beniwal et al., 1992; Martinez et al., 2000) de que o modelo infinitesimal não retrata exatamente o genoma de nenhuma espécie, apesar de ser bastante útil na avaliação genética.

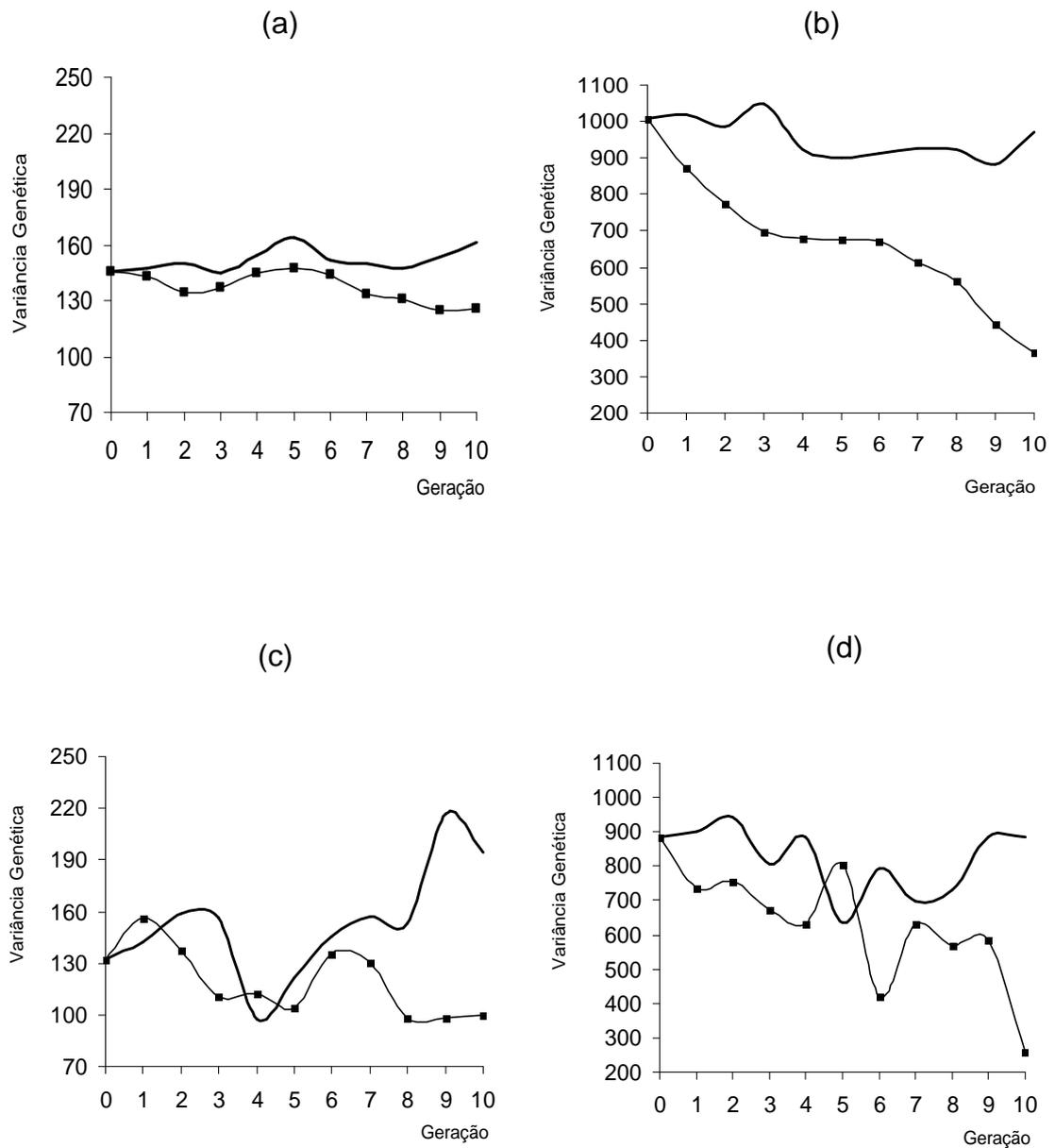


Figura 3 - Valores reais dos componentes de variância genética da população-base e ao longo de 10 gerações, considerando Seleção ao Acaso (—) e Seleção Fenotípica (■) para: (a) população GRAN 900; (b) população GRAN 10; (c) população PEQ 900 e (d) população PEQ 10.

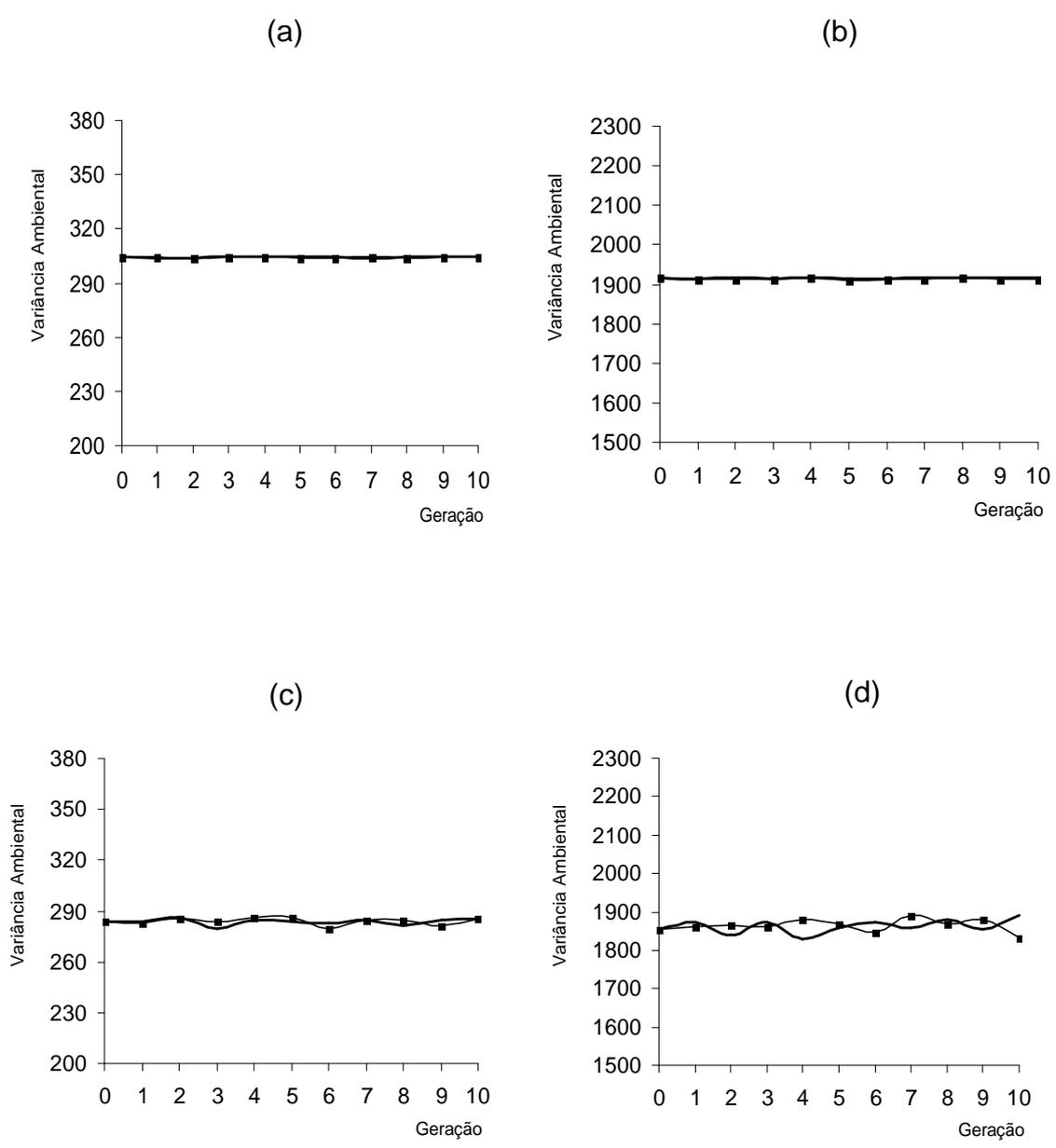


Figura 4 - Valores reais dos componentes de variância ambiental da população-base e ao longo de 10 gerações, considerando Seleção ao Acaso (—) e Seleção Fenotípica (■) para: (a) população GRAN 900; (b) população GRAN 10; (c) população PEQ 900 e (d) população PEQ 10.

Em relação à variância ambiental (Figura 4), nota-se que esta permaneceu constante para todas as populações e para ambos os tipos de seleção, apesar de oscilar um pouco mais no caso das populações pequenas. Esse componente de variância, portanto, parece não sofrer influência da seleção, mesmo quando o número de locos é bastante reduzido.

4.2. Estimação dos Componentes de Variância

As estimativas dos componentes de variância genética aditiva obtidas pelo REML (MTDFREML) e pela metodologia Bayesiana (MTGSAM) para população GRAN 900, para os dois tipos de seleção, encontram-se no Quadro 4. São apresentados também os valores reais das variâncias para a população-base (G_0) e para a última geração após seleção (G_{10}). As estimativas obtidas para G_0 foram relativamente próximas do valor real para ambas as metodologias, sendo as médias da metodologia Bayesiana com *prior* não-informativo e *prior* pouco informativo as mais acuradas. Portanto, para população grande não-endogâmica, que não sofreu seleção e com muitos genes governando a característica, ambas as metodologias produziram resultados bastante satisfatórios.

Na prática, é comum a análise de rebanhos selecionados em que se desconhece os registros dos parentes (ancestrais) do indivíduo, assim como seu próprio parentesco. No caso em que a estimação do componente de variância genética da população-base foi realizada somente com os dados da G_{10} , sem informação de parentesco até a população-base, pode-se verificar que para Seleção ao Acaso as estimativas obtidas pelo REML e pela metodologia Bayesiana, utilizando *priors* não-informativo e pouco informativo, foram bastante próximas entre si, porém superestimadas. A estimativa mais acurada foi a moda obtida pela metodologia Bayesiana com *prior* informativo. Resultados semelhantes foram encontrados quando se realizou Seleção Fenotípica, sendo que as estimativas obtidas pela média e moda da metodologia Bayesiana com *prior* informativo se aproximaram mais do valor real da G_0 .

Quadro 4 – Estimativas dos componentes de variância genética aditiva na população GRAN 900 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			B-NI		B-PI		B-I	
			Média	Moda	Média	Moda	Média	Moda
G₀	146,0	140,3	146,1	153,0	146,0	142,4	148,1	148,6
G₁₀ (SAA)	161,2	203,5	211,1	204,4	206,6	196,0	175,5	169,3
G₁₀ (SF)	125,9	160,3	167,9	160,0	164,2	166,6	153,0	150,4
G₀₋₁₀ (SAA)		152,7	156,1	158,2	156,0	155,8	155,7	154,7
G₀₋₁₀ (SF)		154,8	156,4	155,0	156,4	156,1	156,3	157,3

Ao incluir todos os registros e toda a informação de parentesco da G_{10} até a população-base, verifica-se que para ambos os tipos de seleção e para ambas as metodologias, as estimativas foram próximas umas das outras, porém levemente superestimadas. A semelhança encontrada para ambos os tipos de seleção, associado ao pequeno erro produzido, indicam que para população grande com muitos locos, a Seleção Fenotípica não causa redução na variância genética devido às mudanças na frequência gênica, assim como se espera que ocorra com a Seleção ao Acaso, conforme confirmado pela Figura 3a. Assim, as pressuposições do modelo infinitesimal estão sendo atendidas, exceto pelo fato de haver ligação entre locos na população-base, possível causa do erro encontrado. Dentro desse contexto, os resultados apresentados são satisfatórios. Pode-se observar também que o nível de informação *a priori* não influenciou os resultados, o que pode ser explicado pelo fato de se estar utilizando elevada quantidade de dados (26.400) nessas análises e, portanto, a informação contida na função de verossimilhança “domina” a informação *a priori* (Wang et al., 1994a).

As estimativas dos componentes de variância genética aditiva obtidas pelo REML (MTDFREML) e pela metodologia Bayesiana (MTGSAM) para população GRAN 10, para os dois tipos de seleção, encontram-se no Quadro 5. São apresentados também os valores reais das variâncias para a população-base (G_0) e para a última geração após seleção (G_{10}). As estimativas obtidas para G_0 pelo REML e pela metodologia Bayesiana foram, em geral, próximas umas das outras, sendo as mais acuradas a média e a moda da metodologia Bayesiana com *prior* informativo. A perda da acurácia, verificada nessa população, ao compará-la com a população GRAN 900, está provavelmente associada ao desvio da normalidade dos valores genéticos da própria G_0 , verificado na população GRAN 10 (Quadro 3 e Figura 2) e causado pelo pequeno número de locos.

Analisando os dados da G_{10} somente, nota-se agora, resultados bastante distintos dos obtidos para GRAN 900. Houve grande diferença entre as estimativas obtidas para Seleção ao Acaso e para Seleção Fenotípica. Para o primeiro caso, as estimativas foram levemente superestimadas para ambas as

Quadro 5 - Estimativas dos componentes de variância genética aditiva na população grande GRAN 10 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			B-NI		B-PI		B-I	
			Média	Moda	Média	Moda	Média	Moda
G₀	1.009,2	1.149,5	1.194,1	1.103,8	1.164,4	1.171,8	1.085,9	1.090,4
G₁₀ (SAA)	972,7	1.059,4	1.096,1	1.046,4	1.089,4	1.122,0	1.044,7	1.008,7
G₁₀ (SF)	365,7	356,1	407,3	377,0	435,4	440,1	691,6	707,1
G₀₋₁₀ (SAA)		976,3	1.000,4	992,1	1.000,0	996,7	999,6	991,1
G₀₋₁₀ (SF)		743,8	752,2	753,4	752,7	750,9	759,4	758,7

metodologias empregadas, com exceção da moda da análise Bayesiana com *prior* informativo, que gerou resultado bastante acurado. Realizando Seleção Fenotípica, porém, as estimativas foram bastante distintas do valor real para todas as análises, inclusive pela metodologia Bayesiana com *prior* informativo.

Como para esta população a característica é governada por poucos genes, o decréscimo na variância aditiva é notável ao se realizar Seleção Fenotípica, o que não ocorre quando a seleção é ao acaso. Portanto, ao realizar a análise de dados em que a seleção atuou de forma expressiva sob característica governada por poucos genes, sem incluir registros e informação de parentesco, obtêm-se estimativas com erros expressivos.

Ao incluir todos os registros e toda a informação de parentesco da G_{10} até a população-base, obteve-se para Seleção ao Acaso estimativas muito próximas umas das outras, sendo estas levemente subestimadas. Porém, quando os dados sofreram Seleção Fenotípica, as estimativas foram bem menos acuradas para todas as análises, as quais produziram resultados muito próximos entre si. Esse resultado é bastante interessante, pois mostra que mesmo em populações grandes, com todos os registros e informação de parentesco até a população-base, a quebra da pressuposição do modelo infinitesimal em que não mais a característica é governada por um infinito número de locos de pequeno efeito (mesmo considerando que esses locos têm somente efeito aditivo e que não estão ligados) é suficiente para produzir estimativas bem distantes do valor real. Novamente, percebe-se que para análises com grande quantidade de dados, a informação *a priori* com alto grau de confiança não é capaz de produzir estimativas não viesadas, pois a informação dos dados proporcionada pela função de verossimilhança se sobrepõe à distribuição *a priori* do componente de variância genética aditiva. Resultados de experimentos com camundongos também mostraram a não adequabilidade do modelo infinitesimal na análise de dados que sofreram seleção (Meyer e Hill, 1991; Beniwal et al., 1992).

As estimativas dos componentes de variância genética aditiva obtidas pelo REML (MTDFREML) e pela metodologia Bayesiana (MTGSAM) para população PEQ 900, para os dois tipos de seleção, encontram-se no Quadro 6.

Quadro 6 - Estimativas dos componentes de variância genética aditiva na população PEQ 900 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			B-NI		B-PI		B-I	
			Média	Moda	Média	Moda	Média	Moda
G₀	132,2	140,4	128,4	55,3	131,3	113,5	133,8	131,7
G₁₀ (SAA)	194,8	376,9	474,4	514,0	249,0	240,6	139,9	137,1
G₁₀ (SF)	99,5	104,7	122,5	26,5	112,1	58,2	132,1	127,2
G₀₋₁₀ (SAA)		143,4	146,8	136,1	144,6	138,1	139,2	131,1
G₀₋₁₀ (SF)		140,0	142,4	143,0	140,5	135,5	138,2	138,2

São apresentados também os valores reais das variâncias para a população-base (G_0) e para a última geração após seleção (G_{10}). Para G_0 , a estimativa REML foi levemente superestimada, enquanto que para a metodologia Bayesiana foram encontrados resultados bastante distintos, conforme o nível de informação considerado. Utilizando *prior* não informativo, a metodologia gerou estimativas bastante distintas entre média e moda, sendo a média bem mais próxima do valor real. A associação da pequena quantidade de dados aliada à falta de informação *a priori* é a possível responsável pela *falha* nessa estimação, uma vez que a informação *a priori* tem maior peso no processo quando se possuem poucos dados. Por outro lado, as médias dos *priors* pouco informativo e informativo e a moda do informativo foram bastante acuradas, destacando a importância do nível de informação *a priori* nessa população.

Ao analisar os resultados da G_{10} , verifica-se que para Seleção ao Acaso as estimativas foram superestimadas, com exceção da metodologia Bayesiana com *prior* informativo, a qual produziu estimativas acuradas. Quando se realizou Seleção Fenotípica, novamente estimativas pouco acuradas foram obtidas, com exceção da metodologia Bayesiana com *prior* informativo. Nota-se inclusive grande diferença entre média e moda da análise Bayesiana para *priors* não-informativo e pouco informativo, mostrando a grande falta de simetria das distribuições *a posteriori* geradas.

As estimativas obtidas ao se incluir todos os registros e o parentesco completo dos indivíduos de G_{10} até G_0 para ambos os tipos de seleção e para todas as análises foram, em geral, levemente superestimadas e bastante próximas entre si, resultado este muito semelhante ao encontrado para a mesma situação na população GRAN 900. Como comentado anteriormente, a Seleção Fenotípica não está causando decréscimo substancial na variância aditiva devido à mudança na frequência gênica, sendo as metodologias adequadas para estimar o componente de variância na população-base, uma vez que corrigem a queda na variância causado pela endogamia e pelo desequilíbrio gamético. Novamente, a possível causa do erro encontrado pode estar relacionada à ocorrência de genes ligados na população-base, não satisfazendo assim todas as pressuposições do modelo infinitesimal. Vale ressaltar que incluindo todos os registros até G_0 , passa-se a ter maior

quantidade de dados (1.320 registros), em relação às situações em que havia somente 120 registros, deixando de ser tão relevante o nível de informação *a priori*.

Pelos resultados encontrados, a metodologia Bayesiana com *prior* informativo fornece estimativas acuradas dos componentes de variância genética quando se possui pequena quantidade de dados, ressaltando que o mesmo não se aplica para grandes volumes de dados.

As estimativas dos componentes de variância genética aditiva obtidas pelo REML (MTDFREML) e pela metodologia Bayesiana (MTGSAM) para população PEQ 10, para os dois tipos de seleção, encontram-se no Quadro 7. São apresentados também os valores reais das variâncias para a população-base (G_0) e para a última geração após seleção (G_{10}). Para G_{10} , a estimativa obtida pelo REML foi altamente subestimada, assim como as obtidas pela Bayesiana com *prior* não informativo, cuja densidade *a posteriori* mostrou-se bastante assimétrica. Com *prior* pouco informativo, a densidade *a posteriori* também foi assimétrica, porém sua média mostrou-se levemente subestimada. Ao utilizar *prior* informativo, as estimativas também foram bastante acuradas. A pequena quantidade de dados e um maior desvio dos valores genéticos da normalidade desta população quando comparados com PEQ 900 (ver curtose e assimetria no Quadro 3) podem ser os responsáveis pelo erro encontrado pela metodologia REML. Para a metodologia Bayesiana, esses fatores não prejudicam a análise quando se tem *priors* informativos.

Ao analisar os dados da G_{10} , considerando Seleção ao Acaso, foram encontradas estimativas pouco acuradas pela metodologia REML, pela Bayesiana com *prior* não informativo (média e moda bastante divergentes) e pela moda Bayesiana com *prior* pouco informativo. A média com *prior* pouco informativo e a média e a moda com *prior* informativo foram as que mais se aproximaram do valor real. Ao realizar Seleção Fenotípica, houve falha na estimação pela metodologia REML, assim como pela Bayesiana com *prior* não-informativo. Nota-se novamente, que o desvio da normalidade associado ao tamanho pequeno da população prejudica substancialmente a estimação pela metodologia REML.

Quadro 7 - Estimativas dos componentes de variância genética aditiva na população PEQ 10 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			B-NI		B-PI		B-I	
			Média	Moda	Média	Moda	Média	Moda
G₀	882,7	410,6	494,0	5,0	824,1	591,8	894,1	865,4
G₁₀ (SAA)	882,6	917,0	1.026,5	171,5	897,4	715,4	891,6	851,7
G₁₀ (SF)	259,7	0,0	68,5	1,7	556,4	440,2	847,4	808,9
G₀₋₁₀ (SAA)		1.231,3	1.252,8	1.202,4	1.215,7	1.196,6	1.006,8	1.015,3
G₀₋₁₀ (SF)		882,2	901,2	899,5	886,3	888,4	888,1	880,7

As estimativas obtidas pela Bayesiana com *prior* pouco informativo foram subestimadas, enquanto as mais próximas do valor real foram geradas pela Bayesiana com *prior* informativo.

Incluindo todos os registros e o parentesco completo dos indivíduos da G_{10} até G_0 , verifica-se que para Seleção ao Acaso as estimativas obtidas foram superestimadas para todas as metodologias e, como ressaltado anteriormente, o maior volume de dados ao considerar todas as gerações, diminuiu relativamente a importância da informação *a priori*. Ao realizar Seleção Fenotípica, encontrou-se, de forma geral, estimativas acuradas, sendo, inclusive altamente acurada, a obtida pelo REML. Esse resultado não era o esperado, uma vez que, apesar de se incluir todo o parentesco e todos os registros, a característica é governada por apenas 10 genes. Possivelmente, a própria estrutura dos dados (relacionada os valores reais dos componentes de variância) está favorecendo a ocorrência desses valores.

As estimativas dos componentes de variância ambiental obtidas pelo REML (MTDFREML) e pela metodologia Bayesiana (MTGSAM) para população GRAN 900, GRAN 10, PEQ 900 e PEQ 10 para os dois tipos de seleção encontram-se nos Quadros 24, 25, 26 e 27, respectivamente, no APÊNDICE. São também apresentados nesses quadros os valores reais das variâncias para a população-base (G_0) e para a última geração após seleção (G_{10}).

4.2.1. Porcentagem de Erro

A Porcentagem de Erro (*PE*) das estimativas dos componentes de variância genética aditiva e ambiental obtidas pela metodologia de máxima verossimilhança restrita (REML) e pela metodologia Bayesiana, considerando G_0 , G_{10} e G_0 a G_{10} , provenientes de Seleção Fenotípica encontram-se nas Figuras 5, 6 e 7. Essas figuras complementam a discussão anterior, permitindo melhor comparação entre as estimativas obtidas. Foram utilizadas as modas *a posteriori* geradas pela metodologia Bayesiana na elaboração das figuras.

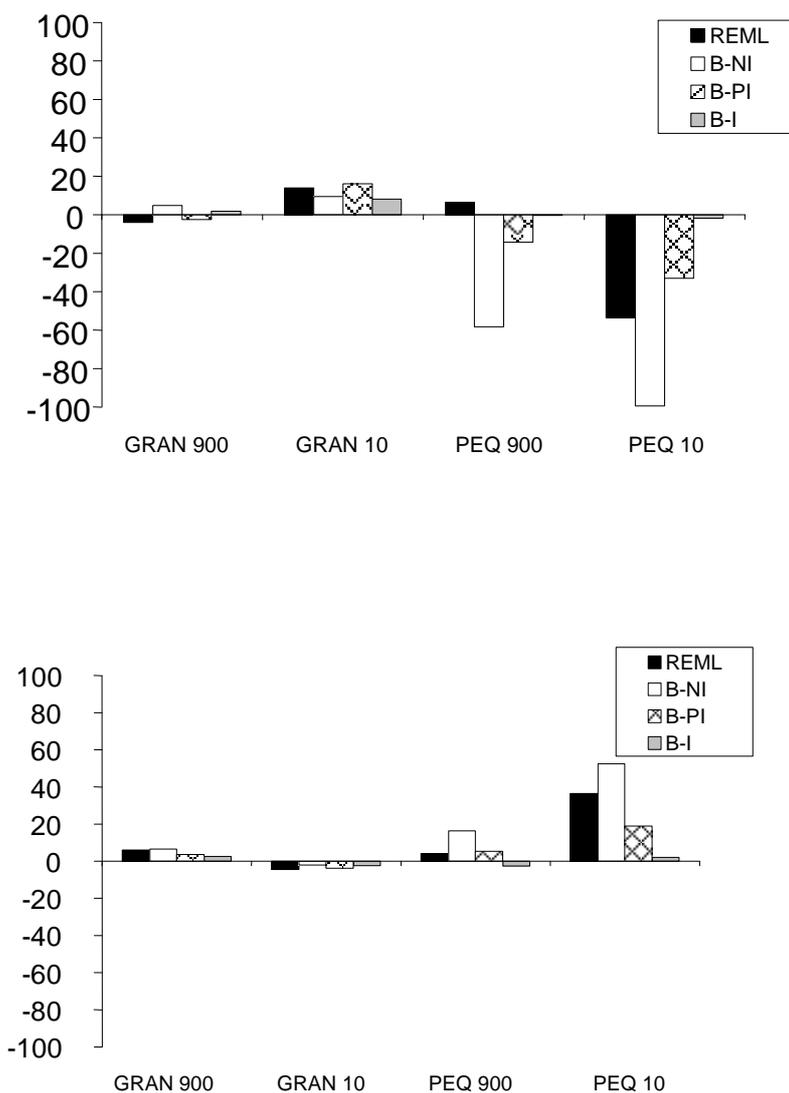


Figura 5 - Porcentagem de Erro das estimativas dos componentes de variância genética aditiva (acima) e ambiental (abaixo) para população da Geração 0 obtidas pela metodologia de máxima verossimilhança restrita (REML) e pela metodologia Bayesiana com três diferentes níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I).

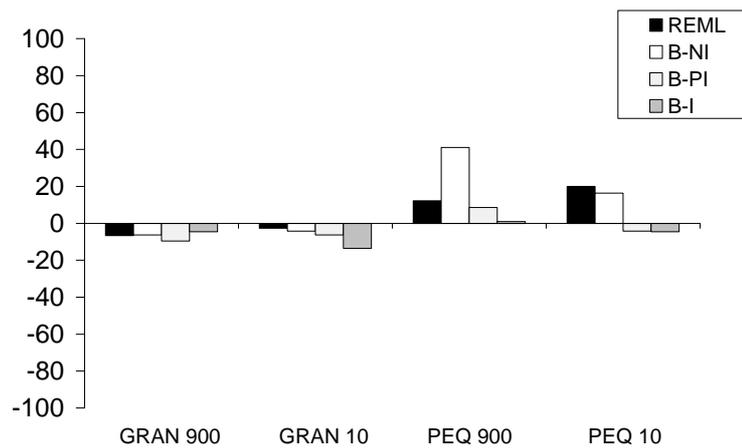
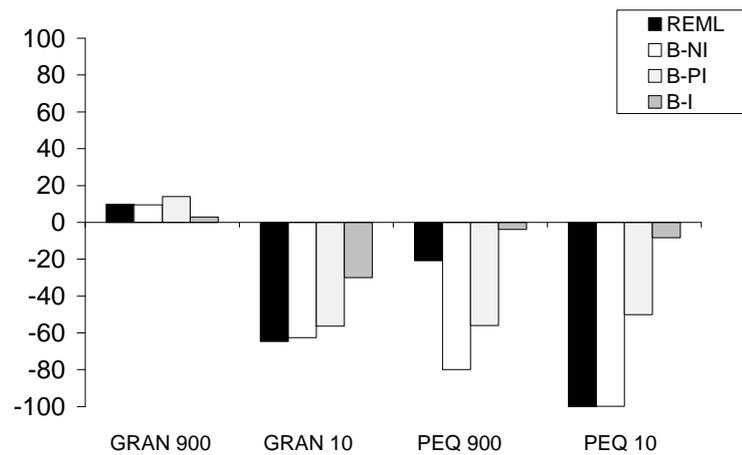


Figura 6 - Porcentagem de Erro das estimativas dos componentes de variância genética aditiva (acima) e ambiental (abaixo) para população da Geração 10 de Seleção Fenotípica, obtidas pela metodologia de máxima verossimilhança restrita (REML) e pela metodologia Bayesiana com três diferentes níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I).

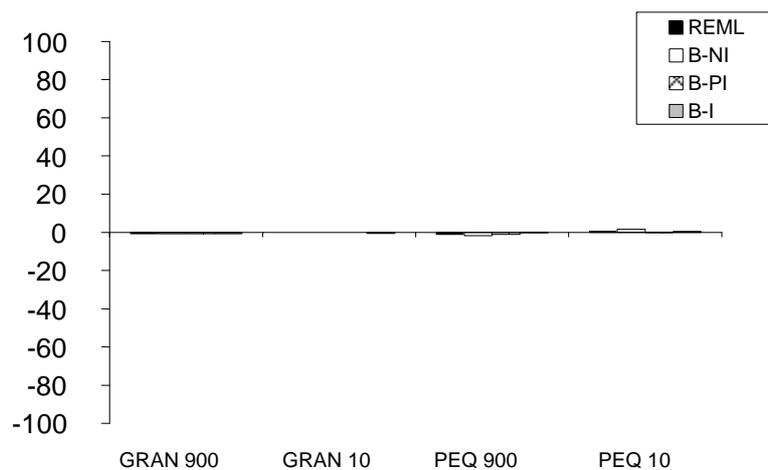
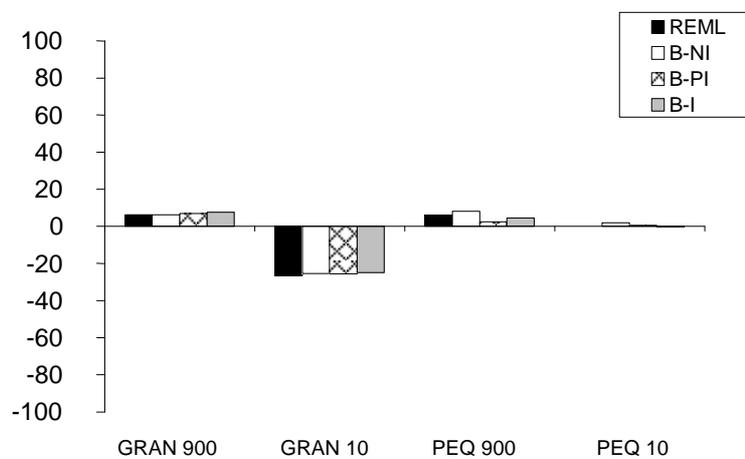


Figura 7 - Porcentagem de Erro das estimativas dos componentes de variância genética aditiva (acima) e ambiental (abaixo) para população formada pelas Gerações 0 a 10 de Seleção Fenotípica, obtidas pela metodologia de máxima verossimilhança restrita (REML) e pela metodologia Bayesiana com três diferentes níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I).

Analisando os resultados produzidos na G_0 (Figura 5), verifica-se que as estimativas obtidas para população GRAN 900 foram bastante acuradas e muito próximas entre si para todas as análises realizadas, tanto para a variância aditiva (cuja maior PE foi de 4,7% para metodologia Bayesiana com *prior* não informativo e a menor foi de 1,8% para metodologia Bayesiana com *prior* informativo) quanto para a ambiental (em que a maior PE foi de 6,6% para metodologia Bayesiana com *prior* não informativo e a menor foi de 2,5% para metodologia Bayesiana com *prior* informativo). Para GRAN 10, as estimativas da variância ambiental foram também próximas entre si para todas as análises e bastante acuradas (o pior resultado foi gerado pelo método REML, sendo a PE igual a -4,5%), enquanto as estimativas da variância genética foram relativamente menos acuradas, em que a PE variou de 8,0% para metodologia Bayesiana com *prior* informativo a 16,1% para metodologia Bayesiana com *prior* pouco informativo. Com relação à população PEQ 900, como comentado anteriormente para variância genética aditiva, houve diferenças mais acentuadas nos valores das estimativas, conforme a metodologia utilizada: a mais acurada foi gerada pela metodologia Bayesiana com *prior* informativo (PE igual a -0,4%) e menos acurada pela metodologia Bayesiana com *prior* não informativo (PE igual a -58,2%). Em relação à variância ambiental, as diferenças entre as metodologias foram de menor magnitude, sendo a menos acurada a estimativa gerada pela análise Bayesiana com *prior* não-informativo (PE igual a 16,5%). Para PEQ 10, as estimativas foram as menos acuradas para ambos os componentes de variância quando comparadas com as demais populações, com exceção das obtidas pela metodologia Bayesiana com *prior* informativo (em que PE foi igual a -2,0% para variância aditiva e igual a 2,2% para variância ambiental). As estimativas menos acuradas foram geradas pela metodologia Bayesiana com *prior* não informativo (PE igual a -99,4%, para variância aditiva e PE igual a 52,6%, para variância ambiental). Com base nesses resultados, percebe-se que mesmo na situação em que a população que está sendo analisada é não-endogâmica e não sofreu seleção, as metodologias utilizadas (com exceção da Bayesiana com *prior* informativo) não forneceram estimativas satisfatórias, quando a característica é governada por poucos genes, principalmente em populações

pequenas. Nota-se também, que para as populações pequenas, para ambos os componentes de variância, as estimativas REML foram relativamente mais acuradas do que as obtidas pela moda da análise Bayesiana com *prior* não-informativo.

Ao observar os resultados obtidos na Figura 6, em que foi considerada somente a G_{10} para se estimar os componentes de variância da população-base, sendo realizada Seleção Fenotípica durante 10 gerações, verifica-se que, para a variância genética aditiva, as PE foram bastante expressivas para todas as populações, com exceção da GRAN 900 e das obtidas pela metodologia Bayesiana com *prior* informativo. Para GRAN 900, a menor e a maior PE foram, respectivamente, de 3,0% para metodologia Bayesiana com *prior* informativo e de 14,1% para metodologia Bayesiana com *prior* pouco informativo. Para GRAN 10, a estimativa menos acurada foi gerada pelo REML (-64,7%) e a mais acurada pela metodologia Bayesiana com *prior* informativo (-29,9%). Em relação à população PEQ 900, o melhor resultado foi obtido para metodologia Bayesiana com *prior* informativo (PE igual a 3,8%) e o pior para metodologia Bayesiana com *prior* não informativo (PE igual a -79,9%). Para a população PEQ 10, as estimativas menos acuradas foram produzidas pelo REML e pela metodologia Bayesiana com *prior* não informativo (PE iguais a -100,0% e -99,9%, respectivamente) e a mais acurada foi obtida pela metodologia Bayesiana com *prior* informativo (PE igual a -8,4%). As estimativas da variância ambiental foram mais acuradas do que as da variância aditiva para todas as populações. Para GRAN 900 e GRAN 10, as menos acuradas foram obtidas pela metodologia Bayesiana com *prior* pouco informativo (PE igual a -9,5%) e pela metodologia Bayesiana com *prior* informativo (PE igual a -13,6%), respectivamente. Para as populações PEQ 900 e PEQ 10, as menos acuradas foram geradas pela metodologia Bayesiana com *prior* não informativo (PE igual a 41,1%) e pelo REML (PE igual a 20,0%), respectivamente.

Realizando Seleção Fenotípica e considerando toda a informação de parentesco e todos os registros da G_0 até G_{10} (Figura 7), verifica-se que os componentes de variância genética aditiva foram estimados satisfatoriamente por todas as metodologias e para todas as populações, com exceção da GRAN 10. Para esta população, todas as metodologias

produziram PE semelhantes, variando de -24,8% a -26,3% para metodologia Bayesiana com *prior* informativo e para o método REML, respectivamente. Em relação à variância ambiental, nota-se que as PE foram insignificantes para todas as populações e metodologias empregadas, variando de 0,0% para população GRAN 10 a 1,6% para população PEQ 10. Nota-se, neste caso, que o tipo de seleção, o número de pares de locos envolvidos na expressão da característica, o tamanho da população, assim como a metodologia utilizada têm influência desprezível na estimação do componente de variância ambiental quando todos os registros e toda a informação de parentesco são incluídos na análise (apesar do desvio da normalidade apresentado para PEQ 10 no Quadro 3). O mesmo não pode ser concluído para variância genética aditiva, conforme discutido anteriormente.

4.2.2. Comparação Entre Média e Moda *a posteriori*

Encontra-se na Figura 8 a Porcentagem de Erro das estimativas da média e da moda *a posteriori* dos componentes de variância genética aditiva (à esquerda) e ambiental (à direita), obtidas pela Metodologia Bayesiana para população da Geração 10 proveniente de Seleção Fenotípica, considerando três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I). Nota-se que, para *prior* não-informativo, a maior diferença ocorreu para a população PEQ 900, em que a média apresentou-se mais acurada, tanto para a variância aditiva quanto para a ambiental. Pequenas diferenças entre as duas foram encontradas nas demais populações para este nível de informação.

Observando-se os resultados para *prior* pouco informativo, nota-se que a maior diferença entre as PE ocorreu na população PEQ 900 para variância genética, novamente sendo a média mais acurada. Para os demais casos, incluindo todas as populações com relação à variância ambiental, as diferenças foram pequenas.

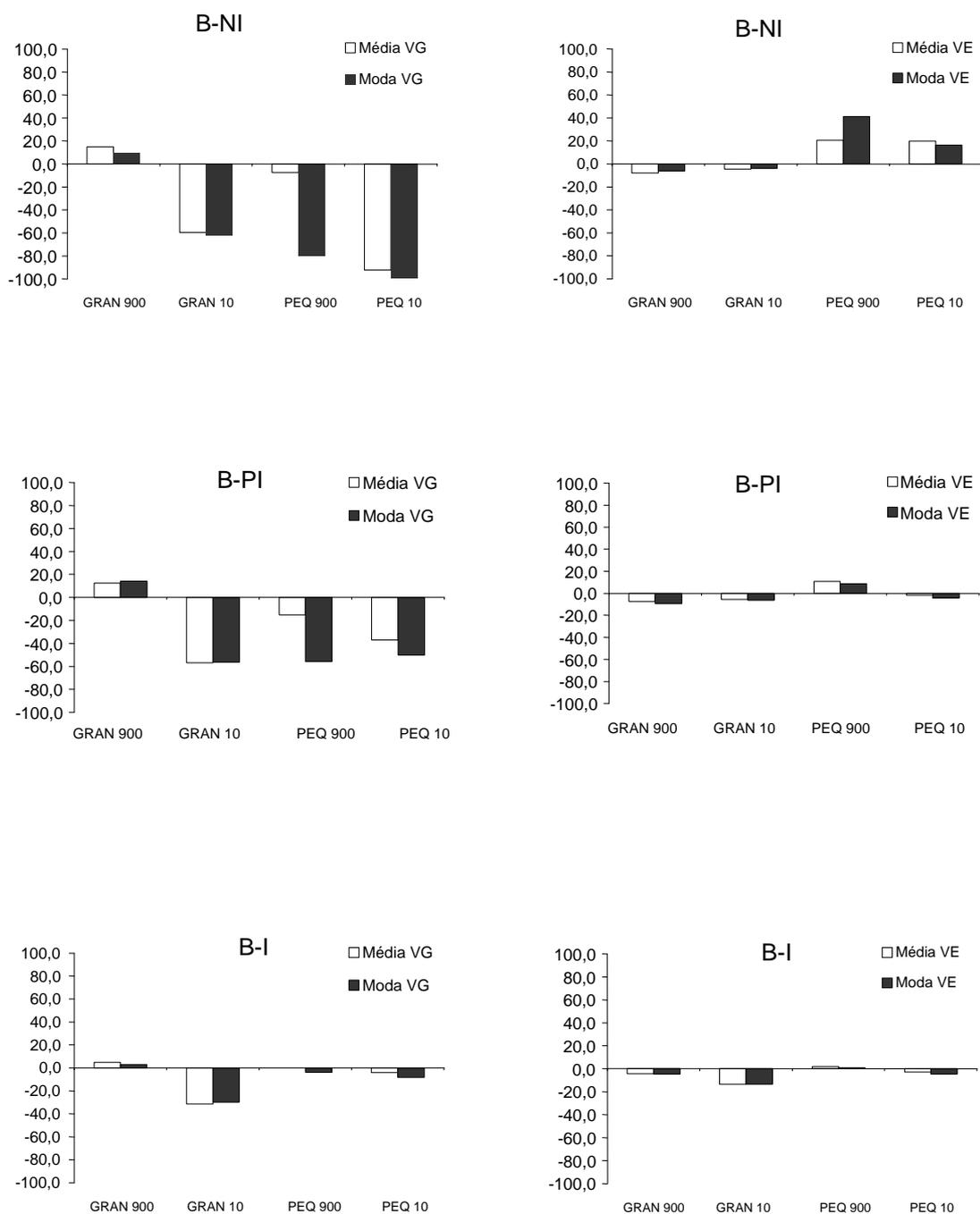


Figura 8 - Porcentagem de Erro da média e da moda *a posteriori* dos componentes de variância genética aditiva (VG, à esquerda) e ambiental (VE, à direita), estimados pela Metodologia Bayesiana para população da Geração 10 proveniente de Seleção Fenotípica, considerando três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I).

Para *prior* informativo, não houve praticamente nenhuma diferença entre as PE entre média e moda Bayesianas, para ambos os componentes de variância.

Foram também verificadas as diferenças entre a PE da média e moda da população-base estimadas ao se incluir a informação de todos os registros e o parentesco completo dos indivíduos da Geração 10 até a população-base. As diferenças encontradas tanto para variância genética aditiva quanto para variância ambiental foram insignificantes, uma vez que as distribuições *a posteriori* foram bastante simétricas.

Pode-se dizer que, em geral, as diferenças mais marcantes ocorreram para populações pequenas, principalmente para a variância genética aditiva, quando o nível de informação foi menor (*priors* não-informativo e pouco informativo).

Schenkel (1998) também estimou componentes de variância pela metodologia Bayesiana e comparou os resultados obtidos pela média e moda *a posteriori*. O autor verificou que as modas *a posteriori* da variância genética aditiva foram menores que as médias *a posteriori* e, geralmente, subestimaram as reais variâncias, uma vez que as distribuições foram assimétricas, principalmente para baixa herdabilidade. Para a variância ambiental, o autor encontrou estimativas das médias e modas *a posteriori* bastante similares, uma vez que as distribuições marginais *a posteriori* desse componente foram aproximadamente simétricas. Segundo Van Tassell et al. (1995), como não existe nenhuma vantagem computacional em se calcular a moda da distribuição *a posteriori* (que, na verdade, é mais difícil de ser calculada), a média é um estimador dos componentes de variância mais apropriado do que a moda.

4.2.3. Adequabilidade do Modelo Infinitesimal

As análises realizadas pelo REML assumem implicitamente uma distribuição normal multivariada dos dados, o que, geneticamente, é equivalente à pressuposição do modelo infinitesimal (Bulmer, 1971). Por este modelo, considera-se que as características são determinadas por infinitos genes aditivos, não ligados, de pequeno efeito, e pressupõe-se que

as frequências gênicas não são alteradas pela seleção. Conseqüentemente, espera-se que as variâncias genéticas permaneçam constantes, exceto como um resultado da endogamia e da seleção, causando desequilíbrio gamético de ligação entre os locos não ligados. Como a metodologia REML considera esses fatores, todas as estimativas da variância genética aditiva provenientes das análises, considerando o acréscimo sucessivo de gerações, devem ser idênticas entre si e iguais à da população-base, caso o modelo infinitesimal seja adequado (Meyer e Hill, 1991).

Encontram-se nas Figuras 9 e 10 as estimativas das variâncias genéticas aditivas obtidas pelo REML com níveis crescentes de inclusão de gerações, para as populações com 900 e 10 genes, respectivamente. Nota-se que na população GRAN 900 (Figura 9), a inclusão das informações de cada nova geração praticamente não modificou o valor das estimativas, ou seja, elas permaneceram constantes para todas as análises. Esse resultado indica que o modelo infinitesimal é adequado para o conjunto de dados em questão. Porém, sabe-se que a estimativa REML do componente de variância, considerando toda a informação da G_{10} até a G_0 , foi levemente superestimada (6% acima do valor real), o que pressupõe-se ter ocorrido devido ao fato de existirem genes ligados na população-base. Portanto, essa estratégia para verificar a adequabilidade do modelo infinitesimal deve ser feita com cautela.

Para a população PEQ 900 (Figura 9), as estimativas da variância genética aditiva permaneceram constantes até a inclusão da quarta geração. Ao incluir a quinta, porém, houve um leve salto de sua estimativa (de 113 para 133), que voltou a se estabilizar até a inclusão da décima geração. Esse leve acréscimo pode estar relacionado com a menor quantidade de dados disponível inicialmente, de forma que, ao aumentar o volume de dados, também se obtiveram estimativas mais acuradas (a PE passou de –14%, em média, para 8%, em média). Portanto, sugere-se que o modelo infinitesimal é apropriado para os dados em questão, incluindo-se aqui a mesma ressalva feita para GRAN 900: sabe-se que a estimativa REML do componente de variância, considerando toda a informação da G_{10} até a G_0 ,

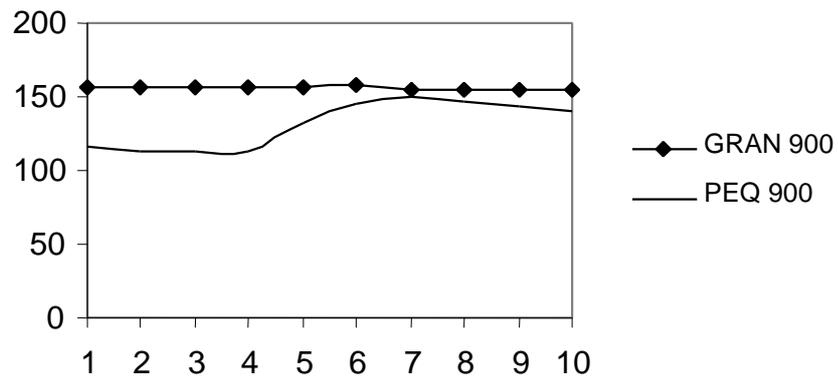


Figura 9 - Estimativas da variância genética aditiva obtidas pelo REML conforme o acréscimo sucessivo de gerações (1:gerações 0 a 1; 2:gerações 0 a 2; 3:gerações 0 a 3; 4:gerações 0 a 4; 5:gerações 0 a 5; 6:gerações 0 a 6; 7:gerações 0 a 7; 8:gerações 0 a 8; 9:gerações 0 a 9; 10:gerações 0 a 10), para as populações GRAN 900 e PEQ 900, com Seleção Fenotípica.

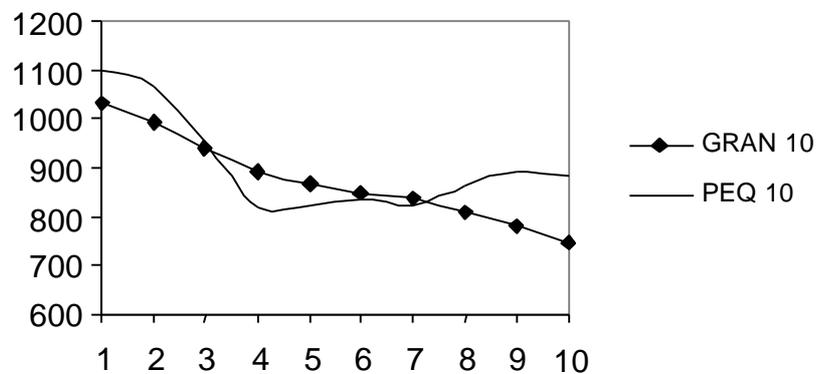


Figura 10 - Estimativas da variância genética aditiva obtidas pelo REML conforme o acréscimo sucessivo de gerações (1:gerações 0 a 1; 2:gerações 0 a 2; 3:gerações 0 a 3; 4:gerações 0 a 4; 5:gerações 0 a 5; 6:gerações 0 a 6; 7:gerações 0 a 7; 8:gerações 0 a 8; 9:gerações 0 a 9; 10:gerações 0 a 10), para as populações GRAN 10 e PEQ 10 com Seleção Fenotípica.

foi também levemente superestimada, sendo a PE a mesma verificada para GRAN 900 (6%), o que pressupõe-se ter ocorrido devido à existência de genes ligados na população-base.

Observando a Figura 10, verifica-se que houve decréscimo marcante das estimativas das variâncias genéticas aditivas para ambas as populações. Para a população GRAN 10, essa queda nos valores das estimativas foi ininterrupta, contrastando com a invariância encontrada para GRAN 900. De acordo com a estratégia adotada, conclui-se que o modelo infinitesimal não é adequado para este conjunto de dados. Este resultado confirma o esperado, uma vez que se observou decréscimo acentuado dos valores reais da variância genética aditiva para população GRAN 10 (Figura 3b) quando comparado com o da população GRAN 900 (Figura 3a).

Para a população PEQ 10, verifica-se que a queda foi intensa até a inclusão da Geração 4, e posteriormente, as estimativas se estabilizaram, havendo leve tendência de acréscimo. Para este conjunto de dados e de acordo com a estratégia de análise adotada, conclui-se que o modelo infinitesimal também é inadequado.

As populações pequenas não se comportaram exatamente como o esperado, conforme ocorreu para as populações grandes, principalmente a PEQ 10. Nota-se que tanto para PEQ 900 quanto para PEQ 10, os valores reais das variâncias genéticas aditivas, apesar de apresentarem tendência de queda, oscilaram bastante entre determinadas gerações (Figura 3c e Figura 3d). Uma possível causa da estabilização das estimativas em PEQ 10 (Figura 10) pode estar relacionada com o acréscimo acentuado da variância aditiva real encontrado nas Gerações 5 (passou de 633 para 805), 7 (passou de 418 para 629) e 9 (passou de 569 para 586), conforme Figura 3d. A estabilização iniciou-se justamente quando se acrescentou as informações da Geração 5.

A estratégia empregada para verificar se o modelo infinitesimal é adequado foi bastante eficiente para identificar a inadequabilidade do modelo devido ao pequeno número de locos, principalmente em populações grandes. Por outro lado, identificou como adequado o modelo infinitesimal em situações em que existem genes ligados na população-base, embora o erro produzido no processo de estimação tenha sido de baixa magnitude.

Assim, a estratégia parece ser eficiente quando se verifica a inadequabilidade do modelo pelo decréscimo das estimativas conforme o acréscimo sucessivo de gerações; porém, quando estas se mostram constantes, deve-se ter cautela ao concluir que o modelo infinitesimal é adequado, pois mesmo assim podem ser obtidas estimativas levemente viesadas.

Diversos trabalhos vêm mostrando que o modelo infinitesimal não é adequado na análise genética de populações sob seleção para determinadas características. Meyer e Hill (1991) realizaram experimento com camundongos, objetivando examinar a extensão do uso do modelo animal na análise de dados sob seleção em longo prazo, utilizando registros e informação de parentesco parciais, para acessar mudanças na variância aditiva, conforme o avanço das gerações. Para ambas as características estudadas (consumo e peso corporal), as estimativas da variância genética aditiva decresceram conforme os dados de gerações mais adiantadas foram incorporados. Os autores sugeriram que o modelo infinitesimal não seria apropriado e que as variâncias teriam mudado como consequência de mudanças nas frequências gênicas devido à seleção. Comentaram também que a metodologia de modelos mistos, em particular o método REML usando um modelo animal, parece ser bastante útil para melhor compreender resultados de experimentos de seleção.

Beniwal et al. (1992) também estimaram as variâncias genéticas em linhas selecionadas de camundongos para peso magro, utilizando o REML com um modelo animal, onde diferentes variâncias foram ajustadas para as linhas individuais ou para os diferentes períodos do experimento. Os autores verificaram redução da variância genética ao longo do experimento, numa magnitude que não poderia ser causada pelo *efeito Bulmer* nem pela endogamia, porque o método REML com o modelo animal estaria considerando esses fatores. Os autores concluíram que a pressuposição do modelo infinitesimal não seria adequada, possivelmente pelo fato de existirem genes que influenciam a característica em estudo que não possuem efeitos suficientemente pequenos. Comentam também que o modelo assume que os locos não estão ligados e, portanto, o desequilíbrio

causado pela seleção entre os locos ligados poderia contribuir para as mudanças na variância genética.

Heath et al. (1995) analisaram dados de experimento de seleção em camundongos para a característica peso corporal e também concluíram que o modelo infinitesimal não pôde explicar adequadamente os resultados observados.

Sorensen et al. (2001) propuseram um método Bayesiano para inferir sobre a variância dentro de um grupo, utilizando toda a informação disponível, sendo este grupo representado, no contexto genético, por indivíduos que pertencem a uma determinada geração. Portanto, o método pode ser utilizado para estudar o comportamento da variância genética durante um experimento de seleção. Conforme comentado pelos autores, esse método difere dos propostos por Sorensen e Kennedy (1984b) e dos propostos por Beniwal et al. (1992) e por Heath et al. (1995) em diversos aspectos.

4.2.4. Estimação em Populações Avançadas

Quando a seleção ocorre por diversas gerações, pode ser de interesse do pesquisador estimar não somente a variância da população-base, mas também a variância em gerações subseqüentes. Segundo Sorensen e Kennedy (1984b) e Sorensen et al. (2001), não está claro como usar todos os dados disponíveis para obter uma estimativa da variância em gerações mais avançadas. Sorensen e Kennedy (1984b) utilizaram dados das gerações 7 e 8, com o intuito de estimar a variância genética aditiva da geração 7 pelo método MIVQUE (considerando seus indivíduos não correlacionados) e obtiveram resultados satisfatórios (o viés encontrado foi de 1%). Segundo os autores, essa metodologia parece adequada para se estimar a variância genética aditiva da geração t , utilizando-se dados da geração t , $t+1$, $t+2$, etc. É importante que todos os indivíduos da geração t , antes de se proceder à seleção, sejam incluídos na análise. Van der Werf e Boer (1990), no entanto, encontraram resultados divergentes dos apresentados por esses autores.

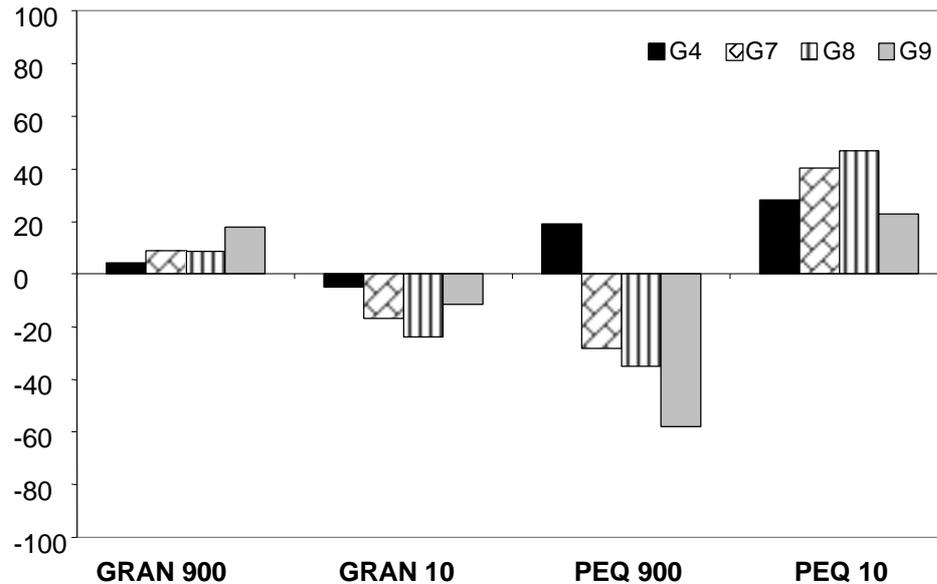


Figura 11 - Porcentagem de Erro das estimativas dos componentes de variância genética aditiva em gerações avançadas, considerando diferentes populações-base (G4: gerações 4 a 10; G7: gerações 7 a 10; G8: gerações 8 a 10; G9: gerações 9 a 10), com Seleção Fenotípica.

A PE das estimativas dos componentes de variância genética aditiva obtidas pelo REML, utilizando somente dados de gerações mais avançadas encontra-se na Figura 11. Para cada situação, consideraram-se diferentes populações como sendo a população-base, ou seja, para cada situação a população-base passou a ser aquela mais antiga. Para a população GRAN 900, a PE tendeu a aumentar (passou de 4,2% para 17,7%) conforme se diminuiu o número de gerações incluídas na análise, afastando-se, cada vez mais da G_0 . Nas três primeiras situações, a PE foi relativamente baixa, o que já não ocorreu quando somente as Gerações 9 e 10 foram analisadas.

Para população GRAN 10, a PE utilizando dados da G_4 até a G_{10} foi pequena (-5,0%), sendo nos demais casos mais expressiva, atingindo o valor de -24,9% quando foram considerados dados da G_8 a G_{10} . Para as populações PEQ 900 e PEQ 10 as estimativas foram pouco acuradas, em que a PE variou de 18,9% a -57,8% para PEQ 900 e de 23,0% a 47,0% para PEQ 10.

De forma geral, esses resultados não são concordantes com os encontrados por Sorensen e Kennedy (1984b), conforme comentado anteriormente.

4.3. Predição dos Valores Genéticos

4.3.1. Quadrado Médio do Desvio

O Quadrado Médio do Desvio (QMD) dos valores genéticos preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*, em relação aos valores genéticos reais, para as populações GRAN 900, GRAN 10, PEQ 900 e PEQ 10 encontram-se nos Quadros 8, 9, 10 e 11, respectivamente.

Verifica-se para todas as populações que praticamente não houve diferença entre os resultados gerados pelo EBLUP, pela análise Bayesiana com *prior* não-informativo e pela análise Bayesiana com *prior* informativo. Para as populações PEQ 900 e PEQ 10, porém, a análise Bayesiana com *prior* informativo gerou QMD menores nas situações G_{10} SAA e G_0 , respectivamente.

Nota-se também que, para todas as populações, os valores mais elevados se referem à situação em que ocorreu Seleção Fenotípica, sendo todos os registros e a informação de parentesco completa dos indivíduos utilizados para predizer os valores genéticos (G_{10} SF 0-10). Esse resultado, provavelmente, ocorreu pelo fato dos valores genéticos preditos nesta situação serem, em média, superestimados, o que pode ser observado na Figura 12, à esquerda para as populações GRAN 900 e GRAN 10 e na Figura 13, à esquerda para as populações PEQ 900 e PEQ 10 (o erro foi sempre crescente, conforme o avanço das gerações). Ao comparar com a situação em que ocorreu Seleção ao Acaso, percebe-se que, em média, os resultados encontrados foram não-viesados (Figura 12, à direita e Figura 13, à direita). Confirma-se no Quadro 12 a alta magnitude do viés para a situação G_{10} SF (0-10), inclusive quando comparada à situação G_{10} SAA (0-10).

Quadro 8 - QMD dos valores genéticos preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I), para GRAN 900

Situação	EBLUP	B-NI	B-I
G₀	87,16	87,00	86,99
G₁₀ SAA	92,33	93,20	89,25
G₁₀ SF	77,57	78,21	76,84
G₁₀ SAA (0-10)	82,81	82,43	82,77
G₁₀ SF (0-10)	3.464,87	3.477,02	3.467,26

Quadro 9 - QMD dos valores genéticos preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I), para GRAN 10

Situação	EBLUP	B-NI	B-I
G₀	556,74	559,20	553,39
G₁₀ SAA	551,18	552,86	550,70
G₁₀ SF	286,33	288,13	326,26
G₁₀ SAA (0-10)	507,66	507,18	508,01
G₁₀ SF (0-10)	18.044,13	18.078,13	18.073,15

Quadro 10 - QMD dos valores genéticos preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I), para PEQ 900

Situação	EBLUP	B-NI	B-I
G₀	89,85	89,03	89,62
G₁₀ SAA	151,32	205,16	88,86
G₁₀ SF	71,22	71,33	72,52
G₁₀ SAA (0-10)	86,54	85,73	85,94
G₁₀ SF (0-10)	2.780,91	2.779,08	2.762,61

Quadro 11 - QMD dos valores genéticos preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I), para PEQ 10

Situação	EBLUP	B-NI	B-I
G₀	723,49	706,31	646,88
G₁₀ SAA	500,14	539,99	500,54
G₁₀ SF	259,62	244,23	275,33
G₁₀ SAA (0-10)	370,94	371,88	362,18
G₁₀ SF (0-10)	17.853,46	17.845,64	17.781,50

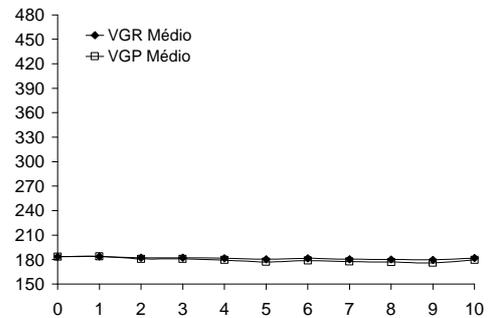
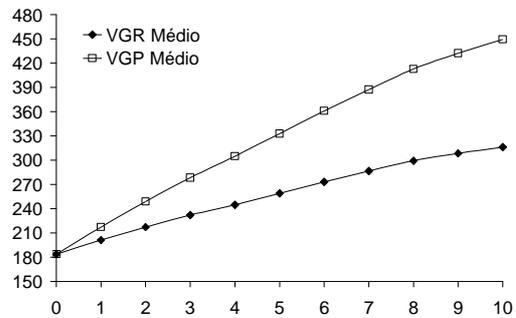
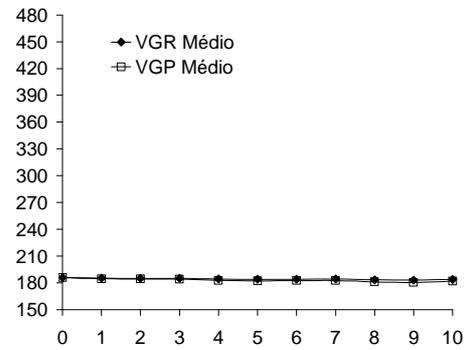
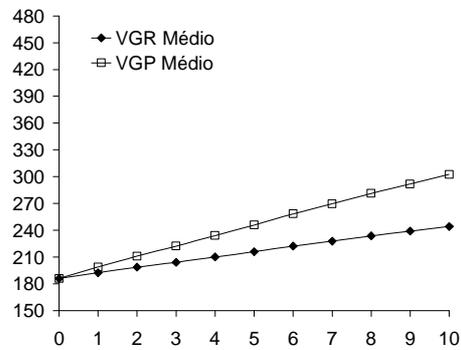


Figura 12 - Valores genéticos médios reais (VGR Médio) e preditos (VGP Médio) pelo EBLUP a cada geração de Seleção Fenotípica (à esquerda) e de Seleção ao Acaso (à direita), considerando todos os registros e parentesco completo dos indivíduos da Geração 10 até a Geração 0 das populações GRAN 900 (acima) e GRAN 10 (abaixo).

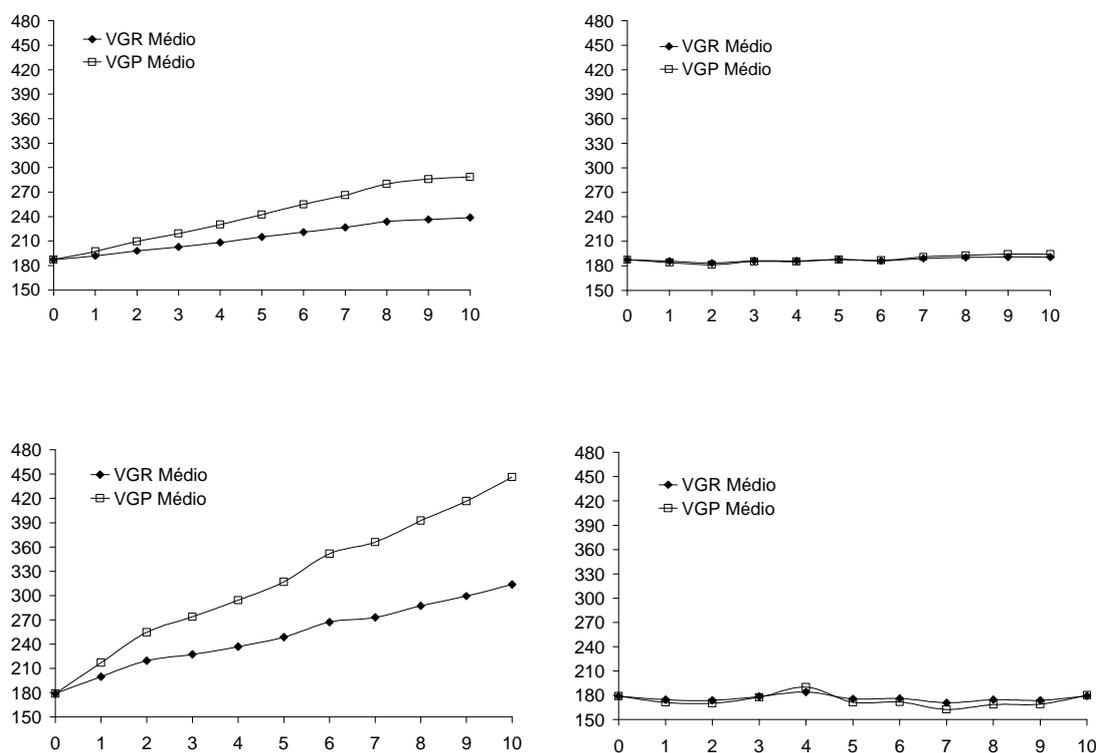


Figura 13 - Valores genéticos médios reais (VGR Médio) e preditos (VGP Médio) pelo EBLUP a cada geração de Seleção Fenotípica (à esquerda) e de Seleção ao Acaso (à direita), considerando todos os registros e parentesco completo dos indivíduos da Geração 10 até a Geração 0 das populações PEQ 900 (acima) e PEQ 10 (abaixo).

Quadro 12 - Desvio médio dos valores genéticos preditos pelo EBLUP em relação aos valores genéticos reais

Situação	GRAN 900	GRAN 10	PEQ 900	PEQ 10
G₀	0,00	0,00	0,00	0,00
G₁₀ SAA	0,00	0,00	0,00	0,00
G₁₀ SF	0,00	0,00	0,00	0,00
G₁₀ SAA (0-10)	-1,83	-2,17	3,53	0,75
G₁₀ SF (0-10)	58,27	133,28	52,12	132,73

Quadro 13 - Equações de regressão linear simples obtidas para os valores genéticos reais (vgr) e preditos pelo EBLUP (vgp) em relação às gerações (G₀ a G₁₀) sob Seleção Fenotípica para os quatro tipos de populações

População	Tendência genética real	Tendência genética estimada
GRAN 900	vgr = 186,6 + 5,8 ** geração	vgp = 187,2 + 11,7 ** geração
GRAN 10	vgr = 189,4 + 13,4 ** geração	vgp = 193,7 + 26,9 ** geração
PEQ 900	vgr = 187,2 + 5,5 ** geração	vgp = 186,9 + 11,0 ** geração
PEQ 10	vgr = 186,6 + 12,7 ** geração	vgp = 192,1 + 25,4 ** geração

** - P<0,01

Os valores genéticos superestimados também influenciam a estimação da tendência genética. Ao regredir os valores genéticos preditos pelo EBLUP e os valores genéticos reais em relação à geração, foram obtidas as equações de regressão, com suas respectivas tendências genéticas apresentadas no 0. Verifica-se que as tendências genéticas para todas as populações foram superestimadas, ou seja, os incrementos genéticos desejáveis, a cada geração de seleção, são, na verdade, bem menores do que os estimados (cerca de 50% menores).

Ao comparar os resultados obtidos conforme o número de pares de locos que governa a característica (conforme apresentado do Quadro 8 ao 11), percebe-se que o QMD foi sempre maior para as populações com poucos pares de locos para todas as situações consideradas. Assim, ao violar a pressuposição do modelo infinitesimal de que a característica é determinada por infinitos pares de locos de efeitos aditivos infinitesimalmente pequenos, perde-se consideravelmente a acurácia na predição dos valores genéticos (Quadro 12). O QMD das situações G₀ e G₁₀ SAA foram maiores do que da situação G₁₀ SF para todas as populações, apesar dos desvios médios para essas três situações terem sido iguais a zero (Quadro 12).

4.3.2. Porcentagem de Seleccionados em Comum

Nos Quadros 14, 15, 16 e 17 encontram-se as Porcentagens de Seleccionados em Comum (presentes na classificação baseada nos valores genéticos preditos e na baseada nos valores genéticos reais) ao seleccionar 15% dos melhores indivíduos em diferentes situações. Nota-se que não houve diferença entre as predições obtidas pela metodologia EBLUP, pela Bayesiana para *prior* não-informativo e pela Bayesiana para *prior* informativo para as populações grandes. Observando a população PEQ 900, verifica-se que, para a situação G_{10} SAA, obteve-se maior Porcentagem de Seleccionados em Comum (66,67%) pela metodologia Bayesiana usando *prior* informativo. Observa-se também que para a situação G_0 da população PEQ 10, foram obtidas maiores porcentagens de indivíduos em comum pela metodologia Bayesiana para *prior* não-informativo e para *prior* informativo (27,28%) do que pelo EBLUP (16,67%). Nas demais situações para as populações pequenas, não houve, de forma geral, diferença nos resultados obtidos pelas metodologias utilizadas.

A Porcentagem de Seleccionados em Comum aumentou ao incluir todos os registros e a informação completa de parentesco, tanto para Seleção ao Acaso quanto para Seleção Fenotípica para todas as populações, exceto para duas situações na população PEQ 900: a porcentagem permaneceu igual para Seleção Fenotípica (11,11%) e diminuiu de 66,67% para 61,11% para Seleção ao Acaso pela análise Bayesiana com *prior* informativo. Essa comparação é feita entre as situações G_{10} SF e G_{10} SF (0-10) e entre G_{10} SAA e G_{10} SAA (0-10). Verifica-se, portanto, que apesar do QMD e do desvio médio aumentarem ao incluir todos os registros e toda informação de parentesco na ocorrência de Seleção Fenotípica, a Porcentagem de Seleccionados em Comum também aumentou nessa situação, ou seja, apesar da diferença entre valores reais e preditos ter aumentado, a inclusão de informação permitiu, de forma geral, melhor classificação dos indivíduos.

Quadro 14 – Porcentagem de indivíduos em comum ao selecionar 15% dos melhores indivíduos, conforme os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população GRAN 900

Situação	EBLUP	B-NI	B-I
G₀	46,11	46,67	46,39
G₁₀ SAA	48,89	48,89	49,44
G₁₀ SF	48,61	48,89	48,89
G₁₀ SAA (0-10)	51,94	52,22	51,94
G₁₀ SF (0-10)	52,78	53,06	52,78

Quadro 15 – Porcentagem de indivíduos em comum ao selecionar 15% dos melhores indivíduos, conforme os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população GRAN 10.

Situação	EBLUP	B-NI	B-I
G₀	43,61	43,89	43,89
G₁₀ SAA	48,61	49,44	49,17
G₁₀ SF	33,06	33,61	33,06
G₁₀ SAA (0-10)	50,28	50,00	50,00
G₁₀ SF (0-10)	36,94	36,94	36,94

Quadro 16 – Porcentagem de indivíduos em comum ao selecionar 15% dos melhores indivíduos, conforme os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população PEQ 900

Situação	EBLUP	B-NI	B-I
G₀	44,44	44,44	44,44
G₁₀ SAA	50,00	44,44	66,67
G₁₀ SF	11,11	11,11	11,11
G₁₀ SAA (0-10)	66,67	66,67	61,11
G₁₀ SF (0-10)	11,11	11,11	11,11

Quadro 17 – Porcentagem de indivíduos em comum ao selecionar 15% dos melhores indivíduos, conforme os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população PEQ 10

Situação	EBLUP	B-NI	B-I
G₀	16,67	27,78	27,78
G₁₀ SAA	44,44	44,44	44,44
G₁₀ SF	22,22	27,78	22,22
G₁₀ SAA (0-10)	55,56	55,56	55,56
G₁₀ SF (0-10)	44,44	44,44	44,44

Com exceção da população GRAN 900 (cujas porcentagens em comum foram bastante próximas), ao comparar tanto G₁₀ SAA com G₁₀ SF quanto G₁₀ SAA (0-10) com G₁₀ SF (0-10), percebe-se que a porcentagem de indivíduos em comum foi menor na ocorrência de Seleção Fenotípica.

Em relação ao número de locos que governa a característica, nota-se que, para as populações grandes, a Porcentagem de Selecionados em Comum diminuiu ao se analisar dados provenientes da população com 10 pares de locos na ocorrência de Seleção Fenotípica. Pelos resultados da metodologia EBLUP, verifica-se que esse valor passou de 48,61% para 33,06% na situação G₁₀ SF e de 52,78% para 36,94% na situação G₁₀ SF (0-10), ao comparar as populações GRAN 900 e GRAN 10. Para as populações pequenas, ocorreu resultado inverso: ao diminuir o número de locos, a porcentagem de indivíduos em comum aumentou, passando de 11,11% para 22,22% na situação G₁₀ SF e de 11,11% para 44,44% na situação G₀₋₁₀ SF. Esse resultado pode estar ocorrendo pelo fato da inclusão de um único indivíduo na população pequena (120 indivíduos) alterar substancialmente a porcentagem de indivíduos em comum selecionados, enquanto que para populações grandes (2.400 indivíduos) essa porcentagem é menor para cada indivíduo incluído. Portanto, para populações pequenas, esta forma de comparação não seria a mais adequada.

4.3.3. Correlação de Ordem

A Correlação de Ordem entre os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*

encontram-se nos Quadros 18, 19, 20 e 21 para as populações GRAN 900, GRAN 10, PEQ 900 e PEQ 10, respectivamente. Para as populações grandes, as correlações obtidas ao se analisar os dados pelas diferentes metodologias não se alteraram. Para as populações pequenas também não houve diferença, exceto para duas situações: a correlação obtida pela metodologia Bayesiana com *prior* informativo foi maior na situação G_{10} SAA para população PEQ 900 e a obtida pelo EBLUP na situação G_0 para população PEQ 10 foi menor entre as demais.

A Correlação de Ordem aumentou com a inclusão de todos os registros e informação completa de parentesco para todas as populações, tanto para Seleção ao Acaso quanto para Seleção Fenotípica. A mesma observação feita anteriormente em relação à Porcentagem de Seleccionados em Comum pode ser realizada: apesar do Quadrado Médio do Desvio e do desvio médio terem aumentado com a inclusão de informação, a Correlação de Ordem aumentou também nessa situação, ou seja, mesmo sendo maior a diferença entre valores reais e preditos, o ordenamento dos indivíduos é melhor realizado quando se consideram todos os registros e informação completa de parentesco.

Com exceção da população GRAN 900 (cujas correlações foram próximas), ao comparar tanto G_{10} SAA com G_{10} SF quanto G_{10} SAA (0-10) com G_{10} SF (0-10), percebe-se que a correlação foi menor na ocorrência de Seleção Fenotípica.

Em relação ao número de locos que governa a característica, nota-se que a Correlação de Ordem diminuiu ao se analisar dados provenientes das populações com poucos locos na ocorrência de Seleção Fenotípica, tanto para a população grande quanto para a pequena. Pelos resultados da metodologia EBLUP, verifica-se que esse valor passou de 0,62 para 0,44 na situação G_{10} SF e de 0,65 para 0,51 na situação G_{10} SF (0-10), ao comparar as populações GRAN 900 e GRAN 10.

Para as populações pequenas, a correlação passou de 0,52 para 0,34 na situação G_{10} SF e de 0,59 para 0,46 na situação G_{10} SF (0-10), ao comparar as populações PEQ 900 e PEQ 10.

Quadro 18 – Correlação de Ordem entre os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população GRAN 900

Situação	EBLUP	B-NI	B-I
G0	0,61	0,61	0,61
G 10 SAA	0,65	0,65	0,65
G 10 SF	0,62	0,62	0,62
G 10 SAA (0-10)	0,69	0,69	0,69
G 10 SF (0-10)	0,65	0,65	0,65

Quadro 19 – Correlação de Ordem entre os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população GRAN 10

Situação	EBLUP	B-NI	B-I
G0	0,66	0,66	0,66
G 10 SAA	0,64	0,64	0,64
G 10 SF	0,44	0,44	0,43
G 10 SAA (0-10)	0,68	0,68	0,68
G 10 SF (0-10)	0,51	0,51	0,51

Quadro 20 – Correlação de Ordem entre os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população PEQ 900

Situação	EBLUP	B-NI	B-I
G0	0,55	0,56	0,55
G 10 SAA	0,69	0,66	0,75
G 10 SF	0,52	0,53	0,52
G 10 SAA (0-10)	0,81	0,81	0,81
G 10 SF (0-10)	0,59	0,59	0,59

Quadro 21 – Correlação de Ordem entre os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação *a priori*: não-informativo (B-NI) e informativo (B-I) para população PEQ 10

Situação	EBLUP	B-NI	B-I
G0	0,49	0,55	0,53
G 10 SAA	0,69	0,63	0,69
G 10 SF	0,34	0,35	0,35
G 10 SAA (0-10)	0,75	0,75	0,76
G 10 SF (0-10)	0,46	0,46	0,46

Com base nos resultados obtidos a partir dos valores genéticos preditos, nota-se que, de forma geral, a análise Bayesiana com *prior* informativo não gerou predições mais acuradas nem mesmo trouxe melhorias no ordenamento dos melhores indivíduos. O mesmo não foi observado na estimação dos componentes de variância, em que, principalmente nas análises com pequena quantidade de dados, o aumento da informação *a priori* foi capaz de produzir estimativas bastante acuradas. Portanto, a predição dos valores genéticos não sofreu influência expressiva das diferentes estimativas dos componentes de variâncias para o mesmo tipo de análise. Esse resultado pode ser explicado pelo fato de se ter realizado análises unicaracterísticas visando prever valores genéticos, que são relativamente insensíveis às mudanças nos valores dos componentes de variâncias (Kennedy, 1991; Slinger, 1975 citado por Schenkel, 1998).

Schenkel et al. (2002) compararam a estimação de efeitos fixos e valores genéticos entre o método convencional EBLUP e a Análise Bayesiana *via* Amostragem de Gibbs. Os autores consideraram combinações de dois modelos de simulação (com e sem efeito fixo de grupo contemporâneo), três tipos de seleção (aleatória, fenotípica e baseada no BLUP), dois níveis de herdabilidade (0,20 e 0,50) e dois níveis de informação de pedigree (0% e 15%, perdida aleatoriamente). As estimativas do efeito fixo de grupo contemporâneo e dos valores genéticos foram, em todas as situações, essencialmente as mesmas com respeito à correlação de ordem. O viés e o quadrado médio do desvio apresentaram o mesmo padrão nas diversas combinações de simulação. Os métodos não foram viesados pela seleção fenotípica nem pela baseada no BLUP quando a informação de pedigree foi completa, embora o quadrado médio do erro dos valores genéticos preditos tenha aumentado para as situações em que o efeito de grupo contemporâneo estava presente. A predição dos valores genéticos pela Análise Bayesiana e pelo EBLUP foram similarmente afetadas pelo efeito conjunto da seleção fenotípica ou baseada no BLUP e da informação de pedigree perdida aleatoriamente. Para ambos os métodos, o viés e o quadrado médio do erro dos valores genéticos e efeitos de grupo contemporâneo estimados aumentaram substancialmente com o transcorrer das gerações. O autor concluiu, portanto, que ambos os métodos mostraram

a mesma habilidade para ordenar os indivíduos e efeitos de grupo contemporâneo.

4.3.4. Avaliação Considerando Diferentes Populações-Base

O QMD entre valores genéticos reais e preditos pelo EBLUP para situações em que se consideraram populações de diferentes gerações como sendo a população-base na ocorrência de Seleção Fenotípica encontram-se no Quadro 22. Conforme se reduziu a quantidade de informação, o QMD também diminuiu para todas as populações. A Seleção Fenotípica adiciona um viés às predições, que não é considerado pela metodologia EBLUP e que pode estar relacionado às violações feitas nas pressuposições do modelo infinitesimal: a existência de genes ligados desde a população-base nas populações GRAN 900 e PEQ 900 e o pequeno número de genes governando a característica nas populações GRAN 10 e PEQ 10.

Encontra-se no Quadro 23 a Correlação de Ordem entre valores genéticos reais e preditos pelo EBLUP, considerando como referência diferentes populações-base. Verifica-se que, para a população GRAN 900 não houve diferença na correlação de ordem, exceto quando se considerou como população-base a própria G_{10} , somente com os registros dos indivíduos desta última geração e parentesco limitado à informação gerada pela inclusão dos pais desses indivíduos (houve leve queda no valor da correlação). Esse resultado praticamente também se repetiu para população GRAN 10, apesar da queda na correlação para G_{10} SF ter sido maior. Nota-se também que, quando comparada com GRAN 900, todas as correlações obtidas para GRAN 10 foram de menor magnitude.

Para a população PEQ 900 praticamente não houve diferença entre as situações G_{10} SF (0-10) e G_{10} SF (4-10), havendo posteriormente tendência de queda na correlação conforme se diminuiu a quantidade de informação. Quando comparada com GRAN 900, as correlações apresentaram menor magnitude.

Quadro 22 - Quadrado Médio do Desvio entre valores genéticos reais e preditos pelo EBLUP, considerando como referência diferentes populações-base

Situação	GRAN 900	GRAN 10	PEQ 900	PEQ 10
G₁₀ SF (0-10)	3.464,87	18.044,13	2.780,91	17.853,46
G₁₀ SF (4-10)	1.242,92	5.468,46	944,99	5.990,29
G₁₀ SF (7-10)	341,12	1.131,57	189,23	2.007,84
G₁₀ SF (8-10)	179,92	536,82	87,69	1.060,34
G₁₀ SF (9-10)	101,62	323,06	79,22	533,14
G₁₀ SF	77,57	286,33	71,22	259,62

Quadro 23 - Correlação de Ordem entre valores genéticos reais e preditos pelo EBLUP, considerando como referência diferentes populações-base

Situação	GRAN 900	GRAN 10	PEQ 900	PEQ 10
G₁₀ SF (0-10)	0,65	0,51	0,59	0,46
G₁₀ SF (4-10)	0,65	0,52	0,58	0,44
G₁₀ SF (7-10)	0,65	0,52	0,54	0,43
G₁₀ SF (8-10)	0,65	0,52	0,54	0,43
G₁₀ SF (9-10)	0,65	0,51	0,51	0,42
G₁₀ SF	0,62	0,44	0,52	0,34

Para população PEQ 10, o maior valor encontrado foi para G₁₀ SF (0-10), havendo queda na correlação conforme se diminuiu a quantidade de informação, sendo o menor valor observado, como nas demais populações, para a situação G₁₀ SF. As menores magnitudes de correlação foram obtidas para PEQ 10, enquanto as maiores, para GRAN 900. Assim como ocorreu para GRAN 10 em relação à GRAN 900, as correlações foram reduzidas para todas as situações para a população PEQ 10, ao serem comparadas com PEQ 900.

5. CONCLUSÃO

Quando a característica é governada por elevado número de genes, os componentes de variância genética aditiva e ambiental são satisfatoriamente estimados em populações selecionadas grandes ou pequenas pelas metodologias usuais, desde que os registros de todos os indivíduos e a matriz completa de parentesco sejam conhecidos. Em populações grandes, no entanto, a exclusão dessas informações não altera substancialmente a acurácia das estimativas.

Quando a característica é governada por reduzido número de genes, estimativas menos acuradas do componente de variância genética aditiva são obtidas em populações grandes, mesmo quando os registros de todos os indivíduos e a matriz completa de parentesco são incluídos nas análises. Caso essas informações sejam desconhecidas, o erro na estimação desse componente aumenta consideravelmente, em populações grandes ou pequenas.

Os valores genéticos são superestimados sob seleção fenotípica quando os registros de todos os indivíduos e a matriz completa de parentesco são incluídos nas análises, independentemente do tamanho da população. A queda na acurácia é ainda mais acentuada quando a característica é governada por reduzido número de genes.

Quando a característica é governada por reduzido número de genes, a classificação correta dos indivíduos com base em seus valores genéticos é prejudicada sob seleção fenotípica, em populações grandes ou pequenas. A inclusão do registro de todos os indivíduos, assim como da matriz de parentesco completa beneficiam a classificação adequada dos indivíduos.

O modelo infinitesimal não é adequado para ser utilizado como pressuposição nas análises genéticas quando a característica é governada por poucos genes, independentemente do tamanho da população.

A estratégia usualmente empregada em experimentos de seleção para verificar a adequabilidade do modelo infinitesimal é eficiente para identificar sua inadequabilidade causada pelo reduzido número de genes. Porém, identifica o modelo infinitesimal como adequado quando há genes ligados na população-base.

As metodologias REML e Bayesiana, em geral, produzem resultados bastante semelhantes na estimação dos componentes de variância. Para análises com menor quantidade de dados, no entanto, estimativas mais acuradas são obtidas ao se utilizar *priors* informativos por meio da análise Bayesiana.

A inclusão de informações externas altamente confiáveis por meio da análise Bayesiana é capaz de gerar estimativas acuradas dos componentes de variância em situações que o método REML não tem sua convergência garantida.

A acurácia na predição dos valores genéticos, assim como a classificação dos indivíduos não são alteradas pelo nível de informação *a priori* das análises Bayesianas, cujos resultados também se assemelham aos da metodologia EBLUP.

REFERÊNCIAS BIBLIOGRÁFICAS

- Barton, N.H.; Keightley, P.D. (2002). Understanding quantitative genetic variation. *Nature Rev. Genet.*, 3:11-21.
- Beniwal, B.K.; Hastings, I.M.; Thompson, R.; Hill, W.G. (1992). Estimation of changes in genetic parameters in selected lines of mice using REML with an animal model. 1. Lean mass. *Heredity*, 69: 352-360.
- Bink, M.C.A.M. (2002). On flexible finite polygenic models for multiple-trait evaluation. *Genet. Res. Camb.*, 80:245-256.
- Blasco, A. (2001). The Bayesian controversy in animal breeding. *J. Anim. Sci.*, 79: 2023-2046.
- Boer, I.J.M.; Van Arendonk, J.A.M. (1992). Prediction of additive and dominance effects in selected or unselected populations with inbreeding. *Theor. Appl. Genet.*, 84:451-459.
- Boldman, K.G., Kriese, L.A., Van Vleck, L.D., Kachman, S.D. (1995). *A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances*. United States Department of Agriculture, Agriculture Research Service.
- Breda, F.C. (2003). *Influência de populações selecionadas sobre a avaliação genética animal*. Viçosa, MG, UFV/MG, Tese (Mestrado em Genética e Melhoramento) - Universidade Federal de Viçosa. 59p.

- Bulmer, M.G. (1971). The effect of selection on genetic variability. *The American Naturalist*, 105(943):201-211.
- Bulmer, M.G. (1976). The effect of selection on genetic variability: a simulation study. *Genet. Res. Camb.*, 28:101-117.
- Casella, G., George, E.I. (1992). Explaining the Gibbs Sampler. *Technical Report No. BU-1098-MA*, Cornell University.
- Dachs, J.N.W. (1988). *Estatística computacional: uma introdução ao Turbo Pascal*. Rio de Janeiro: Livros Técnicos e Científicos, 236p.
- Dempfle, L. (1977). Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs bayésiens. *Ann. Génét. Sél. Anim.*, 9(1):27-32.
- Euclides, R. F. (1996). *Uso do sistema para simulação Genesys na avaliação de métodos de seleção clássicos e associados a marcadores moleculares*. Viçosa, MG, UFV/MG, Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa. 150p.
- Euclides, R.F.; Guimarães, S.E.F. (1997). Associação dos métodos tradicionais de seleção à seleção assistida por marcadores moleculares. *Rev. Bras. Reprod. Anim.*, Belo Horizonte, 21(3):89-96.
- Falconer, D.S.; Mackay, T.F.C. (1996). *Introduction to quantitative genetics*. Longman, England, 4 ed., 464 p.
- Fernando, L.R.; Stricker, C.; Elston, R.C. (1994). The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theor. Appl. Genet.*, 88:573-580.
- Fisher, R. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Trans. Royal Soc. Edinburgh*, 52:399-433.

- Geman, S., Geman, D. (1984). *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721-741.
- Gianola, D., Fernando, R.L. (1986). Bayesian methods in animal breeding theory. *J. Anim. Sci.*, 63: 217-244.
- Gianola, D.; Folley, J.L. (1982). Non linear prediction of latent genetic liability with binary expression: An empirical Bayes approach. In: *World Congr. Genet. Appl. Livest. Prod., 5. Proceedings...*Madrid, Spain, 7: 293-303.
- Gianola, D.; Foulley, J.L. (1990). Variance estimation from integrated likelihoods. *Genet. Sel. Evol.*, 22:403-417.
- Gianola, D.; IM, S.; Fernando, L. (1988). Prediction of breeding value under Henderson's selection model: a revisitation. *J. Dairy. Sci.*, 71(10):2790-2798.
- Goffinet, B. (1983). Selection on selected records. *Genet. Sel. Evol.*, 15:91-98.
- Gonçalves, T.M. (2003). Genes de efeito principal e locos de características quantitativas (QTL) em suínos. Tese de Doutorado, Faculdade de Medicina Veterinária e Zootecnia - UNESP, Campus de Botucatu - SP, 82p.
- Graser, H.U.; Smith, S.P.; Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.*, 64:1362-1370.
- Harville, D. (1974). Bayesian inferences for variance components using only error contrasts. *Biometrika*, 61:383-385.

- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57:97-109.
- Heath, S.C.; Bulfield, G., Thompson, R.; Keightley, P.D. (1995). Rates of change of genetic parameters of body weight in selected mouse lines. *Genet. Res.*, 66:19-25.
- Henderson, C.R. (1973). Sire evaluation and genetic trends. In: *Animal Breeding Genetic Symposium in Honor of Dr. J.L. Lush, 1973, Champaign. Proceedings...* Champaign, ASAS/ADSA, p.10-28.
- Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31: 423-447.
- Henderson, C.R.; Kempthorne, O.; Searle, S.R.; Krosigk. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15 (6): 192-218.
- Hill, W.G. (1999). Advances in quantitative genetics theory. In: *From Jay Lush to Genomics: visions for the future of animal breeding and genetics. Proceedings...* Iowa State University, Ames, Iowa, p. 35-46. [disponível em : <http://www.agbiotechbio.com>]
- Hill, W.G. (2002). Factors influencing the maintenance of variability in genetic improvement programmes. In: *Anais de palestras da XXXIX Reunião Anual da Sociedade Brasileira de Zootecnia*, Recife-PE, p. 226-245.
- Hudson, G.F.S.; Schaeffer, L.R. (1983). Monte Carlo comparison of sire evaluation models in populations subject to selection and nonrandom mating. *J. Dairy Sci.*, 67:1264-1272.
- Juga, J.; Thompson, R. (1989). Estimation of variance components in populations selected over multiple generations. *Acta Agric. Scand.*, 39:79-89.

- Kennedy, B.W. (1991). C.R. Henderson: the unfinished legacy. *J. Dairy Sci.*, 74:4067-4081.
- Littell, R.C., Freund, R.J., Spector, P.C. (1991). *SAS[®] System for linear models*. 3 ed. Cary, NC: SAS Institute Inc., 329p.
- Martinez, V.; Bünger, L.; Hill, W.G. (2000). Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions. *Genet. Sel. Evol.*, 32:3-21.
- Meyer, K.; Hill, W.G. (1991). Mixed model analysis of a selection experiment for food intake in mice. *Genet. Res. Camb.*, 57:71-81.
- Muir, W.M. (1997). Genetic selection strategies: computer modeling. *Poultry Science*, 76:1066-1070.
- Patterson, H.D.; Thompson, R. (1971). Recovery of inter-block information when block size are unequal. *Biometrics*, 58: 545-554.
- Pieramati, C.; Van Vleck, L.D. (1993). Effect of genetic groups on estimates of additive genetic variance. *J. Anim. Sci.*, 71:66-70.
- Pollak, E.J.; Quaas, R.L. (1981). Monte Carlo study of within-herd multiple trait evaluation of beef cattle growth traits. *J. Anim. Sci.*, 52(2):248-256.
- Quaas, R.L.; Pollak, E.J. (1980). Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.*, 51(6):1277-1287.
- Raftery, A.E.; Lewis, S.M. (1992). *How many iterations in the Gibbs sampler?* [disponible em: <http://www.stat.washington.edu/www/research/online>], 14p.

- Robertson, A. (1977). The effect of selection on the estimation of genetic parameters. *Z. Tierz. Züchtbiol*, 94:131-135.
- Ronningen, K. (1971). Some properties of the selection index derived by "Henderson's mixed model method". *Z. Tierz. Züchtbiol*. 83: 186-193.
- Schaeffer, L.R.; Schenkel, F.S.; Fries, L.A. (1998). Selection bias on animal model evaluation. In: *World Congr. Genet. Appl. Livest. Prod.*, 6. *Proceedings...Armidale, Australia*, 25, p.501-508.
- Schenkel, F.S. (1998). *Studies on effects of parental selection on estimation of genetic parameters and breeding values of metric traits*. Ph.D. thesis. University of Guelph, 191p.
- Schenkel, F.S.; Schaeffer, L.R. (1998). Effects of non translation invariant selection on estimates of variance components. In: *World Congr. Genet. Appl. Livest. Prod.*, 6. *Proceedings... Armidale, Australia*, p.509-512.
- Schenkel, F.S., Schaeffer, L.R., Boettcher, P.J. (2002). Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. *Genet. Sel. Evol.*, 34:41-59.
- Slanger, W.D. (1975). *Topics relating to the variance and covariance components needed to evaluate animals genetically*. Ph.D. thesis. Cornell University, Ithaca, New York, USA., 124p.
- Smith, S.P.; Graser, H.U. (1986). Estimating variances components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.*, 69:1156-1165.
- Sorensen, D.; Gianola, D. (2002). *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Springer, USA, 740p.

- Sorensen, D.; Fernando, R; Gianola, D. (2001). Inferring the trajectory of genetic variance in the course of artificial selection. *Genet. Res. Camb.*, 77:83-94.
- Sorensen, D.A.; Johansson, K. (1992). Estimation of direct and correlated responses to selection using univariate animal models. *J. Anim. Sci.* 70:2038-2044.
- Sorensen, D.A., Kennedy, B.W. (1983). The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. *Theor. Appl. Genet.* 66:217.
- Sorensen, D.A., Kennedy, B.W. (1984a). Estimation of response to selection using least squares and mixed model methodology. *J. Anim. Sci.*, 58(5):1097-1106.
- Sorensen, D.A., Kennedy, B.W. (1984b). Estimation of genetic variances from unselected and selected populations. *J. Anim. Sci.*, 59(5):1213-1223.
- Sorensen, D.A.; Kennedy, B.W. (1986). Analysis of selection experiments using mixed model methodology. *J. Anim. Sci.*, 63:245-258.
- Sorensen, D.A., Wang, C.S., Jensen, J. (1994). Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet. Sel. Evol.*, 26:333-360.
- Thompson, R. (1977). The estimation of heritability with unbalanced data II. Data available on more than two generations. *Biometrics*, 33:497-504.
- Thompson, R. (2001). Statistical validation of genetic models. *Livest. Prod. Sci.*, 72:129-134.

- Turelli, M.; Barton, N.H. (1994). Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics*, 138:913-941.
- Van der Werf, J.H.J. (1992). Restricted Maximum Likelihood estimation of additive genetic variance when selected base animals are considered fixed. *J. Anim. Sci.*, 70:1068-1076.
- Van der Werf, J.H.J; Boer, I.J.M. (1990). Estimation of additive genetic variance when base populations are selected. *J. Anim. Sci.* 68:3124-3132.
- Van Kaam, J.B.C.H.M. (1998). "GIBANAL" – *Analyzing program for Markov Chain Monte Carlo Sequences (Version 2.10)*. Manual, Department of Animal Sciences, Wageningen Agricultural University.
- Van Tassel, C.P., Casella, G., Pollak, E.J. (1995). Effects of selection on estimates of variance components using Gibbs Sampling and Restricted Maximum Likelihood. *J. Dairy Sci.*, 78:678-692.
- Van Tassel, C.P., Van Vleck, L.D. (1995). *A manual for use of MTGSAM. A set of Fortran programs to apply Gibbs Sampling to animal models for variance components estimation*. United States Department of Agriculture, Agriculture Research Service, 85p.
- Walsh, B. (2000). *Evolution and selection of quantitative traits*. [disponível em <http://nitro.biosci.arizona.edu>]
- Wang, C.S., Gianola, D., Sorensen, D.A., Jensen, J., Christensen, A., Rutledge, J.J. (1994a). Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis. *Theor. Appl. Genet.*, 88:220-230.
- Wang, C.S., Rutledge, J.J., Gianola, D. (1993). Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet. Sel. Evol.*, 25:41-62.

Wang, C.S., Rutledge, J.J., Gianola, D. (1994b). Bayesian analysis of mixed linear models *via* Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.*, 26: 91-115.

APÊNDICE

Quadro 24 - Estimativas dos componentes de variância ambiental na população GRAN 900 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			B-NI		B-PI		B-I	
			Média	Moda	Média	Moda	Média	Moda
G₀	304,2	322,8	319,6	324,2	320,06	315,45	316,10	311,84
G₁₀ (SAA)	304,3	259,6	255,7	256,6	259,78	264,24	280,5	279,1
G₁₀ (SF)	304,2	284,2	279,7	285,3	281,94	275,14	291,1	290,3
G₀₋₁₀ (SAA)		303,7	302,0	302,1	302,14	301,20	302,5	300,8
G₀₋₁₀ (SF)		301,9	301,5	301,7	301,61	301,66	301,68	302,00

Quadro 25 - Estimativas dos componentes de variância ambiental na população GRAN 10 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			NI		PI		I	
			Média	Moda	Média	Moda	Média	Moda
G₀	1.915,2	1.829,0	1.808,7	1.876,6	1.826,0	1.845,7	1.883,1	1.869,7
G₁₀ (SAA)	1.916,7	1.870,3	1.854,9	1.852,1	1.854,0	1.858,7	1.886,1	1.873,1
G₁₀ (SF)	1.911,3	1.865,9	1.829,0	1.836,5	1.806,0	1.795,3	1.654,8	1.655,6
G₀₋₁₀ (SAA)		1.906,4	1.894,5	1.904,8	1.897,6	1.901,6	1.896,7	1.888,9
G₀₋₁₀ (SF)		1.915,9	1.913,8	1.915,3	1.913,7	1.915,6	1.912,5	1.905,1

Quadro 26 - Estimativas dos componentes de variância ambiental na população PEQ 900 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			NI		PI		I	
			Média	Moda	Média	Moda	Média	Moda
G₀	283,5	295,4	317,9	330,4	302,3	298,9	288,9	276,4
G₁₀ (SAA)	285,1	92,2	40,0	0,3	193,6	175,3	275,9	265,5
G₁₀ (SF)	285,2	318,0	341,9	400,2	314,1	308,0	288,6	286,4
G₀₋₁₀ (SAA)		300,9	300,7	292,4	301,0	298,1	301,2	300,5
G₀₋₁₀ (SF)		280,4	280,8	278,4	280,7	280,5	281,4	282,4

Quadro 27 - Estimativas dos componentes de variância ambiental na população PEQ 10 obtidas pelo REML (MTDFREML) e por metodologia Bayesiana com três níveis de informação *a priori*: não informativo (B-NI), pouco informativo (B-PI) e informativo (B-I), considerando Seleção ao Acaso (SAA) e Seleção Fenotípica (SF)

Geração	Valor Real	REML	Metodologia Bayesiana					
			NI		PI		I	
			Média	Moda	Média	Moda	Média	Moda
G₀	1.853,6	2.530,9	2.606,9	2.827,7	2.167,0	2.203,4	1.933,5	1.894,1
G₁₀ (SAA)	1.889,9	1.964,8	2.003,8	2.482,6	1.960,3	1.967,1	1.887,8	1.860,6
G₁₀ (SF)	1.833,4	2.223,7	2.219,7	2.156,3	1.823,3	1.776,8	1.800,2	1.769,3
G₀₋₁₀ (SAA)		1.728,1	1.727,9	1.730,1	1.736,0	1.737,4	1.835,9	1.818,8
G₀₋₁₀ (SF)		1.860,9	1.863,0	1.883,4	1.859,8	1.849,7	1.859,1	1.864,1