

MARCOS YAMAKI

**ESTIMAÇÃO DE PARÂMETROS GENÉTICOS DE PRODUÇÃO  
DE LEITE E DE GORDURA DA RAÇA PARDO-SUIÇA,  
UTILIZANDO METODOLOGIAS FREQUENTISTA E BAYESIANA**

Dissertação apresentada à  
Universidade Federal de Viçosa, como parte  
das exigências do Programa de Pós-  
Graduação em Zootecnia, para obtenção do  
título de *Magister Scientiae*

VIÇOSA  
MINAS GERAIS - BRASIL  
2006

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

T

Y19e  
2006

Yamaki, Marcos, 1980-

Estimação de parâmetros genéticos de produção de leite e de gordura da raça Pardo-suiça, utilizando metodologias freqüentista e bayesiana / Marcos Yamaki.

– Viçosa : UFV, 2006.

v, 64f. : il. ; 29cm.

Orientador: Robledo de Almeida Torres.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia..

1. Bovino de leite - Genética. 2. Leite - Produção.  
3. Máxima verossimilhança restrita. 4. Teoria bayesiana de decisão estatística. I. Universidade Federal de Viçosa.  
II. Título.

CDD 22.ed. 636.2082

MARCOS YAMAKI

**ESTIMAÇÃO DE PARÂMETROS GENÉTICOS DE  
PRODUÇÃO DE LEITE E DE GORDURA DA RAÇA PARDO-  
SUÍÇA, UTILIZANDO METODOLOGIAS FREQUENTISTA E  
BAYESIANA**

Dissertação apresentada à  
Universidade Federal de Viçosa, como parte  
das exigências do Programa de Pós-  
Graduação em Zootecnia, para obtenção do  
título de *Magister Scientiae*

APROVADA: 31 de julho de 2006.

---

Prof. José Ivo Ribeiro Júnior

---

Prof. Antônio Policarpo de Souza  
Carneiro

---

Prof. Paulo Sávio Lopes  
(Co-orientador)

---

Prof. Ricardo Frederico Euclides  
(Co-orientador)

---

Prof. Robledo de Almeida Torres  
(Orientador)

**À minha família;**

**aos meus amigos;**

**à minha “rosa”;**

**Dedico este trabalho.**

## AGRADECIMENTOS

À Universidade Federal de Viçosa, em especial ao Departamento de Zootecnia, pela oportunidade de realização deste curso.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudo.

Aos meus pais, Shigekazu e Kazue Yamaki, pelo carinho, educação e patrocínio dos meus projetos de vida.

Às minhas irmãs, Karin e Tatiana Yamaki pelo incentivo e convivência.

À minha rosa pela compreensão apesar de muitas vezes compartilhar apenas a companhia da solidão no nosso longínquo planeta B612 enquanto eu desbravava o universo paramétrico da máxima verossimilhança.

Ao meu grande e querido amigo e orientador, Professor Robledo de Almeida Torres, pela amizade, confiança, orientação, críticas e sugestões viabilizando a execução deste trabalho.

Ao Professor Ricardo Frederico Euclides, amigo e futuro orientador no doutorado, pela elaboração do Mitsub especialmente para atender às minhas necessidades, pelas sugestões e ensinamentos.

Ao Professor Paulo Sávio Lopes, pelos valiosos ensinamentos durante todo o curso, pelas sugestões, críticas e amizade.

Ao Professor Cláudio Vieira de Araújo, por disponibilizar os dados, pela paciência, esclarecimentos e sugestões.

À Professora Lúcia Galvão de Albuquerque, pelas rápidas, mas não menos importantes sugestões, pela atenção, por todo suporte que me foi concedido em minha estadia na Unesp-Jaboticabal e que sem a qual, dificultaria muito a realização deste trabalho.

Ao meu grande e querido amigo Lindenberg, pela amizade, ensinamentos, paciência e todo apoio e atenção que me foram dedicados para a realização deste trabalho.

À Joãozinho pela rápida e essencial ajuda nos procedimentos do SAS.

Aos colegas e amigos de curso e do peito, André, Alex, Bundinha, Carla, Débora, Galego, Gilberto, Gustavo, Humberto, Jane, Kécia, Kleibe, Leandro, Luciara, Marília, Marcelinho, Mário, Rodrigo, Rodriguim (e aos que eu possivelmente tenha me esquecido!) pela amizade, companheirismo e pelo convívio agradável e confraternizações regadas à cerveja.

À minha turma de graduação, Zootecnia 2000, por todos os momentos inesquecíveis que passamos e passaremos juntos, pela amizade, cachaça e companheirismo.

Aos amigos Thiago, Daniel e Matheus pelo constante incentivo e eterna amizade.

Ao tio Johnny e tia Rose, pelo carinho, amizade, ensinamentos e todo apoio que sempre me prestaram.

Aos efeitos de ambiente permanente que Viçosa me proporcionou.

Aos meus inimigos, pela oportunidade de crescimento.

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

## ÍNDICE

RESUMO .....	vi
ABSTRACT .....	viii
INTRODUÇÃO GERAL .....	- 1 -
REVISÃO DE LITERATURA .....	- 4 -
1.1. MÁXIMA VEROSSIMILHANÇA RESTRITA .....	- 4 -
1.2. MÉTODOS BAYESIANOS .....	- 7 -
1.3. AVALIAÇÃO GENÉTICA DA RAÇA PARDO-SUIÇA .....	- 15 -
REFERÊNCIAS BIBLIOGRÁFICAS .....	- 17 -
CAPÍTULO 1 .....	- 21 -
ABSTRACT .....	- 23 -
1. INTRODUÇÃO .....	- 25 -
2. MATERIAIS E MÉTODOS .....	- 27 -
3. RESULTADOS E DISCUSSÃO .....	- 30 -
4. CONCLUSÃO .....	- 35 -
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	- 37 -
CAPÍTULO 2 .....	- 39 -
ABSTRACT .....	- 40 -
1. INTRODUÇÃO .....	- 41 -
2. MATERIAIS E MÉTODOS .....	- 43 -
2.1. DESCRIÇÃO DOS DADOS .....	- 43 -
2.2. MODELOS .....	- 44 -
2.3. ANÁLISES .....	- 45 -
3. RESULTADOS E DISCUSSÃO .....	- 47 -
4. CONCLUSÃO .....	- 56 -
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	- 57 -
CONSIDERAÇÕES FINAIS .....	- 59 -

## RESUMO

YAMAKI, Marcos, M. S., Universidade Federal de Viçosa, julho de 2006.  
**Estimação de parâmetros genéticos de produção de leite e gordura da raça Pardo-suíça, utilizando metodologias freqüentista e bayesiana.**  
Orientador: Robledo de Almeida Torres. Co-orientadores: Paulo Sávio Lopes, Ricardo Frederico Euclides

Dados de primeira lactação de 6.262 vacas distribuídas em 311 rebanhos, filhas de 803 touros com partos entre os anos de 1980 e 2003 foram utilizados para estimar de componentes de variância para as características de produção de leite e gordura com informações de primeira lactação, em animais da raça Pardo-Suíça. Os componentes de variância foram estimados pelo método da máxima verossimilhança restrita (REML) e Bayesiano, sob modelo animal, por meio de análises uni e bicaracterística. A estimação realizada via REML foi obtida com o programa MTDFREML (BOLDMAN *et al.* 1995) testando modelos unicaracterística com diferentes efeitos para as covariáveis e considerados grupo contemporâneo e estação como efeitos fixos. Os melhores ajustes obtidos nas análises unicaracterística foram utilizados na análise bicaracterística. A duração da lactação reduziu a estimativa da variância aditiva quando era utilizada no modelo sugerindo que os animais estariam sendo corrigidos para uma mesma base quanto à capacidade de imprimir duração da lactação mais longa ou mais curta à progênie sendo, portanto, não recomendado o ajuste para esta covariável. Já a idade da vaca ao parto, influenciou linearmente a produção de leite e gordura. As herdabilidades estimadas foram 0,26 e 0,25 para produção de leite e gordura respectivamente com correlação genética de 0,95. A alta correlação entre a produção de leite e gordura obtida sugere que parte dos genes que atuam na produção de leite também responde pela produção de gordura, de tal forma que a seleção para a produção de leite resulta, indiretamente, em aumentos na produção de gordura. A estimação via inferência Bayesiana foi realizada com o programa MTGSAM – (VAN TASSELL E VAN VLECK, 1995). Foram testados diversos tamanhos de cadeia para a obtenção das densidades marginais a *posteriori* das análises unicaracterística, a melhor proposta para o tamanho de cadeia, *burn-in* e amostragem foi utilizada para a análise bicaracterística. Os períodos de *burn-in* foram testados pelo programa GIBANAL (VAN KAAM, 1998) cujas análises

fornecem um intervalo de amostragem para cada *burn-in* testado, o critério de escolha do intervalo de amostragem foi feito de acordo com a correlação serial, resultante do *burn-in* e do processo de amostragem. As estimativas de herdabilidade obtidas foram  $0,33 \pm 0,05$  para ambas as características com correlação de 0,95. Resultados similares foram obtidos em estudos utilizando a mesma metodologia em informações de primeira lactação. A fase estacionária foi adequadamente atingida com uma cadeia de 500.000 iterações e descarte inicial de 30.000 iterações.

## ABSTRACT

YAMAKI, Marcos, M. S., Universidade Federal de Viçosa, July 2006.  
**Estimation of genetic parameters of milk and fat yield of Brown-Swiss cows using frequentist and bayesian methodologies.** Adviser: Robledo de Almeida Torres. Co-Advisers: Paulo Sávio Lopes and Ricardo Frederico Euclides

First lactation data of 6.262 Brown-Swiss cows from 311 herds, daughters of 803 sires with calving between 1980 and 2003 were used to estimate genetic parameters for milk and fat production traits. The components of variance were estimated by restricted maximum likelihood (REML) and bayesian methods, using animal model with uni and two-traits analysis. The estimation by REML was obtained with the software MTDFREML (BOLDMAN *et al.* 1995) testing unitrait models with different effects to covariables and considering contemporary group and season as fixed effect. The best fitting obtained on unitrait analysis were used on two-trait analysis. The estimative of additive variance was reduced when lactation length was included on the model suggesting that the animals were been fitted to the same base on the capacity of transmit a longer or shorter lactation length to the progeny. Therefore, fitting to this covariable is not recommended. On the other side, the age of calving has linearly influenced milk and fat production. The heritability estimates were 0,26 and 0,25 to milk and fat yield respectively with genetic correlation of 0,95. the high correlation among these traits suggests that part of genes that acts on milk yield also respond to fat yield, in such way that selection for milk yield results, indirectly, in increase on fat yield. The estimation by Bayesian inference was made on software MTGSAM – (VAN TASSELL E VAN VLECK, 1995). Chain lengths were tested to obtain the marginal posterior densities of unitrait analysis, the best option of chain length, burn-in and sampling interval was used on two-trait analysis. The burn-in periods were tested with the software GIBANAL (VAN KAAM, 1998) witch analysis inform a sampling interval for each burn-in tested, the criteria for choosing the sampling interval was made with the serial correlation resulting by burn-in and sampling process. The heritability estimates were  $0,33 \pm 0,05$  for both traits with genetic correlation of 0,95. Similar results were obtained on studies using the same methodology on first

lactation records. The stationary phase adequately reached with a 500.000 chain length and 30.000 burn-in iterations.

## INTRODUÇÃO GERAL

A pecuária leiteira do Brasil está vivendo novos tempos. A produção cresce, a população não sofre mais crise de desabastecimento, o mercado interno ganhou proteção, a qualidade do leite progrediu e deve melhorar ainda mais, a entressafra está quase extinta e os preços para os produtores ganharam boa estabilidade.

Há três anos atrás, os produtores amargavam preços baixos, deteriorados, enquanto que atualmente os preços estão bem mais estáveis. Também colaborou para o bom momento do setor não apenas as medidas *antidumping*, mas também a conquista do mercado externo pelo leite nacional. Em 2004, pela primeira vez em toda sua história houve superávit na balança comercial de lácteos: as exportações foram maiores do que as importações.

Segundo a FAO, o Brasil produziu 23.320.000 toneladas de leite em 2005 e atualmente ocupa a sétima posição na produção mundial. Hoje é o maior produtor de leite da América Latina e o segundo das Américas, ficando atrás somente dos EUA, maior produtor mundial.

Neste cenário favorável do mercado do leite, o conhecimento do potencial produtivo dos animais destinados à produção leiteira e dos fatores que interferem na expressão deste potencial, sejam ambientais ou genéticos, são primordiais para que sejam traçadas as estratégias da exploração pecuária leiteira.

Os sistemas de produção de leite no Brasil são constituídos, em sua maioria, por animais da raça holandesa e seus cruzamentos com outras raças principalmente de origem zebuínas, como o Gir Leiteiro e o Guzerá. Entretanto novas opções vêm chamando a atenção dos produtores como a utilização de outras raças européias especializadas em produção de leite, tanto de animais

puros como de seus cruzamentos. Dentre estas, a raça Pardo-Suíça, tem atendido às exigências de diversos sistemas de produção.

Segundo a Associação Brasileira de Criadores de Gado Pardo-Suíço, animais da raça são registrados no Brasil desde 1938 e tem apresentado aumento no número de registros especialmente nas últimas décadas o que denota o crescente interesse pela criação de animais desta raça. Hoje são 69.000 animais registrados distribuídos por todos os estados da federação. Quando são contados animais sem registro e com participação em cruzamentos, o número de animais da raça sobe para mais de um milhão de cabeças.

Com tamanha participação no rebanho leiteiro brasileiro, é de extrema importância que os programas de melhoramento genético da raça Pardo-Suíça sejam bem fundamentados com estratégias de avaliação e uso de animais de potencial genético comprovado.

A eficiência dos programas de melhoramento genético depende da acurácia com que os indivíduos submetidos à seleção são avaliados. No melhoramento animal é de extrema importância o conhecimento das propriedades genéticas das populações, ou seja, a avaliação do valor genético com o objetivo de classificar os melhores indivíduos que serão os pais da próxima geração, e, também, quantificar a contribuição destes para o ganho genético.

A decisão sobre quais animais serão utilizados para a reprodução, é em geral baseada na análise de informações coletadas em vários rebanhos que apresentam grande variabilidade de sistemas de produção. Esta peculiaridade da pecuária brasileira torna ainda mais difícil a escolha correta destes animais. Assim, deve-se utilizar a metodologia que melhor prediz o valor genético dos animais.

Existem vários métodos para estimação componentes de variância; todavia, o método da Máxima Verossimilhança Restrita – REML (“Restricted Maximum Likelihood”), desenvolvido por PATTERSON & THOMPSON (1971), é o mais usado atualmente. O REML baseia-se no princípio de maximização do logaritmo da função densidade de probabilidade das observações, que considera a perda de graus de liberdade na estimação dos efeitos fixos.

Outro método que vem sendo utilizado para obtenção dos componentes de (co)variância e para avaliação genética baseia-se na teoria Bayesiana, iniciada no século XVIII. A inferência Bayesiana utiliza métodos probabilísticos para descrever a incerteza sobre o verdadeiro valor de algum parâmetro (BLASCO, 2001), considerando-se a existência de conhecimento, ou desconhecimento, inicial a respeito deste.

O objetivo deste trabalho foi estimar componentes de variância para produção de leite e gordura de animais da raça Pardo-Suíça aplicando modelos uni e bicaracterística via REML. Utilizar as estimativas obtidas via REML nos valores iniciais dos processos iterativos das análises uni e bicaracterística por inferência Bayesiana, determinar o tamanho de cadeia e o intervalo de amostragem ideal para estimação dos componentes de variância para as características em estudo da população em questão.

# REVISÃO DE LITERATURA

## 1.1. Máxima Verossimilhança Restrita

O conhecimento prévio dos componentes de (co)variância é necessário na predição dos valores genéticos, quando se faz uso de métodos de predição como o BLUP (Best Linear Unbiased Predictor). Entretanto, estes componentes não são geralmente conhecidos e podem ser estimados por vários métodos.

O método da Máxima Verossimilhança (ML), derivado por HARTLEY e RAO (1967), citados por THOMPSON (1979), utiliza-se de formas quadráticas calculadas, levando-se em consideração os próprios valores dos parâmetros. Portanto, as esperanças dessas formas quadráticas não são lineares nos parâmetros e devem ser solucionadas iterativamente, o que dificulta a obtenção das estimativas. Por outro lado, o uso dessas formas quadráticas mais complexas é efetivo, já que, segundo ROTHSCILD *et al.* (1979), elas fornecem estimativas não-viesadas dos componentes de variância na presença de seleção, o que não ocorre com o Método I de Henderson.

Para obtenção dos estimadores por ML, é necessário o conhecimento da distribuição da variável em estudo. De acordo com ANDERSON (1984), este método é iterativo, fornece estimativas não-negativas dos componentes de variância e elimina o “viés” atribuído às mudanças nas frequências gênicas resultantes da seleção, porém não considera a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo. A estimação por meio do método ML requer a inversa da matriz da parte de efeitos aleatórios das equações de modelos mistos, e isto pode ser computacionalmente difícil, ou aproximadamente impossível, para esquemas complexos que envolvem vários fatores aleatórios com números grandes de níveis e por isso, segundo BLASCO (2001), o ML não foi utilizado para determinação do valor genético até recentemente. Outra limitação do ML é de que a normalidade é requerida para

que os estimadores tenham a propriedade de máxima verossimilhança (KENNEDY, 1991).

PATTERSON & THOMPSON (1971) desenvolveram o método que é atualmente utilizado, denominado método da Máxima Verossimilhança Restrita (REML). A diferença deste para o método ML é que, ao invés de se utilizar toda a função de verossimilhança, se utilizam apenas dos termos da verossimilhança relativos a contrastes ortogonais da parte aleatória das observações, pois os contrastes entre os efeitos fixos fornecem nenhuma informação adicional sobre o resíduo, ou sobre os efeitos aleatórios. O método REML tem a vantagem de dar o mesmo resultado da ANOVA em dados balanceados (ANOVA é não viesada com variância mínima para tal situação).

De acordo com MEYER (1986) o método da máxima verossimilhança restrita (REML) é o mais recomendado para dados desbalanceados. A mesma autora considera o método do REML mais adequado para estimação de parâmetros genéticos dos dados de melhoramento animal, pois além de considerar a perda de graus de liberdade resultante da estimação dos efeitos fixos, as estimativas caem sempre dentro do espaço paramétrico. Esse método evita erros de pequenas amostras associadas aos efeitos fixos e dados provenientes de populações submetidas à seleção. O REML e ML têm a notável propriedade na estimação de parâmetros da população base livre de viés devido à seleção (HENDERSON, 1986).

Existem alguns autores que questionam a superioridade do método REML, mas GIANOLA E FERNANDO (1986), utilizando-se de argumentos Bayesianos, apresentam razões consistentes para que o método REML seja considerado superior ao ML. Segundo esses autores, os chamados efeitos fixos são parâmetros de distúrbio, quando os interesses são os componentes de variância dos efeitos aleatórios, e o que se faz, quando se excluem os contrastes dos efeitos fixos e se utilizam apenas a verossimilhança dos contrastes entre as partes aleatórias das observações, é eliminar os parâmetros de distúrbio por meio de integração. Isto equivale a fazer inferências sobre a distribuição marginal dos componentes de (co)variância, o que é coerente, visto que esses componentes são obtidos numa etapa anterior e, portanto, devem ser válidos para quaisquer que sejam os valores dos efeitos fixos a serem estimados.

De modo geral, a obtenção das estimativas pelo REML implica no uso de métodos que demandam grande esforço computacional, em termos de memória e tempo de processamento. Para contornar esse problema GRASER *et al.* (1987) propuseram um algoritmo para análise de uma característica, no qual o ponto de máximo do logaritmo da função densidade de probabilidade, da parte aleatória das observações, é determinado por meio de sucessivas avaliações da função, a partir de valores atribuídos à razão entre os componentes de variância genética aditiva e residual. Este algoritmo não envolve a derivação da função densidade de probabilidade, em relação aos componentes de variância, para o estabelecimento do sistema de equações. Em razão disso, foi denominado DFREML (*Derivative-Free Restricted Maximum Likelihood*). BOLDMAN *et al.* (1995) desenvolveram um aplicativo, o MTDFREML (*Multiple-Trait Derivative-Free Maximum Likelihood*), para características múltiplas, em modelos com dois ou mais fatores aleatórios.

A não-normalidade dos dados, a assimetria e a curtose associadas aos dados também influenciam as inferências obtidas, sendo mais problemáticas em análises multivariadas (COCHRAN & COX, 1978; BROWNIE *et al.*, 1990; e AZZALINI & VALLE, 1996).

Apesar da pressuposição de normalidade dos dados ser crucial para o procedimento de estimação por ML e REML, a literatura não é informativa sobre como o não atendimento desta exigência influencia as estimativas de parâmetros genéticos no melhoramento animal (FREITAS, 2000). Na análise de variância, entretanto, há evidências de que os valores da curtose e, em menor extensão, da assimetria dos dados (SCHEFFÉ, 1959) são os indicadores mais importantes de como os desvios de normalidade dos dados influenciam as inferências obtidas. Segundo COCHRAN & COX (1978), a não-normalidade dos dados pode influir negativamente na estimação dos efeitos fixos, no uso dos testes t e F e na heterogeneidade da variância do erro.

Este fato comprova a hipótese de homogeneidade de variância por meio dos efeitos fixos, assumida na REML, em que o método mais utilizado para estimação de variâncias e covariâncias, nem sempre é atendido. Assumir esta hipótese na estrutura de dispersão dos dados, quando a mesma não é atendida, implica em prejuízos na resposta à seleção dos animais (GIANOLA *et al.*, 1992).

## 1.2. Métodos Bayesianos

Métodos Bayesianos podem ser especialmente avaliados em problemas complexos ou nas situações em que naturalmente não há conformidade com o cenário clássico; muitos problemas genéticos caem dentro dessas categorias.

GIANOLA *et al.* (1982) introduziram os métodos Bayesianos no melhoramento animal no contexto de característica de limiar e posteriormente GIANOLA & FERNANDO (1986) adicionalmente realçou as possibilidades de explorar as técnicas Bayesianas.

As técnicas Bayesianas foram abandonadas no passado porque usualmente requerem complicada resolução de múltiplas integrais, muitas vezes com uso de métodos numéricos (CANTET *et al.*, 1992), que pode não ser flexível. A descoberta da técnica MCMC (Monte Carlo Markov Chain) tem dado uma solução para muitos problemas que não foram resolvidos no passado devido à impossibilidade de resolução dessas integrais. Entretanto, surgiram novos problemas que são relacionados à convergência das cadeias de Gibbs. Felizmente, esses novos problemas são facilmente manuseáveis, particularmente quando a distribuição dos dados é normal (BLASCO, 2001). Detalhes sobre aplicação desta técnica na análise de experimentos de seleção podem ser encontrados em SORENSEN *et al.* (1994).

Os métodos MCMC são um conjunto de processos iterativos para a geração aproximada de amostras de distribuições multivariadas. Neles, 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 o Amostrador de Gibbs (GERMAN E GERMAN, 1984) um caso especial deste algoritmo que vem sendo empregado por diversos autores na avaliação genética animal.

O Amostrador de Gibbs é uma técnica indireta para gerar variáveis aleatórias a partir de uma distribuição marginal, sem, no entanto, a necessidade de calcular a sua densidade (ou seja, a função densidade de probabilidade marginal desta variável). Essa técnica é relativamente fácil de ser implementada. Dada a função de máxima verossimilhança e as densidades *a priori*, calcula-se a densidade conjunta *a posteriori* dos parâmetros desconhecidos. A partir dessa densidade, obtém-se a distribuição condicional completa de cada variável, fixando-se as demais variáveis da densidade conjunta. Esse conjunto de densidades condicionais completas permite a implementação do Amostrador de Gibbs. Obtidas as densidades marginais, pode-se calcular estatísticas de interesse das distribuições *a posteriori*. Todos os cálculos necessários para implementar o Amostrador de Gibbs são feitos com escalares, sem requerer a inversão de matrizes.

A maior dificuldade encontrada atualmente é sobre a detecção da convergência da(s) cadeia(s) de Gibbs. Assim, os estudos sobre inferência Bayesiana têm se voltado para a verificação da convergência de métodos de integração numérica.

Para verificar se a estacionalidade foi atingida, é necessário examinar a(s) cadeia(s) gerada(s) e verificar se ela(s) apresenta(m) certas características. Uma recomendação comum é descartar a parte inicial da cadeia, a qual é chamada de *burn-in* e basear os diagnósticos nos elementos restantes da(s) cadeia(s). Teoricamente uma cadeia de Markov converge para sua distribuição de equilíbrio. Entretanto, na prática, certificar a convergência de uma cadeia, gerada por um Amostrador de Gibbs, não é tarefa fácil (WANG *et al.*, 1994).

Várias formas de monitoramento da convergência são encontradas na literatura. Contudo, infelizmente, esses procedimentos não são totalmente infalíveis (CASELLA & GEORGE, 1992).

Parametrização centrada pobremente dos efeitos aleatórios pode, às vezes, causar baixa convergência na cadeia de Gibbs (GELFAND *et al.*, 1995).

Naturalmente, a definição do número de iterações que devem ser descartadas está relacionada com a velocidade de convergência da cadeia. GELMAN *et al.* (1995) sugerem descartar a primeira metade das iterações; enquanto VAN TASSEL *et al.* (1995) descartaram apenas as 100 primeiras

iterações de uma cadeia de tamanho igual a 5.000. O Amostrador de Gibbs requer o monitoramento do correto período de Burn-in (WOLFINGER & KASS, 2000).

A principal conclusão de vários autores é a dificuldade de saber quando uma cadeia de Gibbs tenha convergido para a distribuição *a posteriori* conjunta.

Na estatística Bayesiana não existe distinção entre variáveis fixas e aleatórias, todas as variáveis são consideradas como aleatórias. Efeitos considerados fixos no melhoramento animal (rebanho, ano, etc) são tidos, na estatística Bayesiana, como variáveis aleatórias sobre as quais se tem pouco ou nenhum conhecimento *a priori*, ou seja, *a priori* para o pesquisador é indiferente ao provável valor destas variáveis (GIANOLA & FERNANDO, 1986).

A interpretação para probabilidades se refere às probabilidades subjetivas, em que é levado em consideração o “grau de confiança” de um determinado indivíduo. As técnicas Bayesianas utilizam probabilidades subjetivas medindo graus de confiança sobre os valores dos parâmetros desconhecidos. Essas probabilidades subjetivas são usadas para definir o que é chamado de distribuição *a priori* para o parâmetro. Então, ao trabalhar com métodos Bayesianos, tem-se o parâmetro desconhecido como uma variável aleatória, cuja distribuição *a priori* (antes de se ter os valores observados da amostra) é conhecida; essa distribuição *a priori* resume o grau de confiança subjetivo sobre o valor desconhecido do parâmetro.

Uma crítica comum da aproximação Bayesiana é que a escolha da distribuição *a priori* é bastante subjetiva. Este questionamento está relacionada com o fato que, em alguns casos, a distribuição *a posteriori* é muito sensível à escolha da distribuição *a priori*. Neste caso, dois pesquisadores usando o mesmo dado podem chegar em diferentes conclusões se usar diferentes distribuições *a priori* (SHOEMARKER, 1999).

Na estatística Bayesiana, o conhecimento *a priori* é mais importante quando as informações disponíveis são escassas ou pouco informativas. Quando se tem grande volume de dados (por exemplo, grande número de progênies por reprodutor) as informações *a priori* tendem a ser subjugadas ou dominadas pela função de verossimilhança. O teorema de Bayes fornece uma solução precisa para cada conjunto de dados, independente do seu tamanho (GIANOLA & FERNANDO, 1986).

A escolha apropriada das distribuições *a priori* ainda é uma questão complicada dentro da inferência Bayesiana (SORENSEN *et al.*, 1994), principalmente quando são assumidos “priors” não-informativos. Essa escolha pode levar a distribuições *a posteriori* impróprias, de difícil reconhecimento, principalmente nos métodos de integração numérica.

Segundo GIANOLA & FERNANDO (1986) a estatística Bayesiana não requer normalidade nem linearidade dos dados. Também, as variâncias não precisam ser conhecidas, elas são estimadas a partir dos dados pelos métodos iterativos. Em relação a dados de populações selecionadas, os autores citam que a distribuição *a posteriori* dos valores genéticos e parâmetros é a mesma com ou sem seleção ou acasalamento controlado. Assim a estatística Bayesiana pode fornecer inferência a respeito dos valores e parâmetros genéticos mesmo na presença de seleção.

Infelizmente, análises mais ricas requerem mais intensivo custo computacional. Com o rápido avanço na tecnologia computacional, é provável que vários modelos sejam manipulados eficientemente num futuro próximo (WANG *et al.*, 1994).

Ambas as escolas, Freqüentista e Bayesiana, utilizam a verossimilhança na realização de suas inferências. Por um lado, a Verossimilhança assume um papel central na inferência Bayesiana como a função que expressa toda a informação proveniente dos dado. Por 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).

### 1.2.1. Teorema de Bayes

Considerando  $\theta$  um vetor de parâmetros a serem estimados, um vetor de observações  $\mathbf{y}$  e a função densidade conjunta  $f(\theta, \mathbf{y})$ , conforme a teoria da probabilidade, tem-se:

$$f(\theta, \mathbf{y}) = f(\mathbf{y}|\theta) f(\theta) \quad [3]$$

e

$$f(\theta, \mathbf{y}) = f(\theta|\mathbf{y}) f(\mathbf{y}) \quad [4]$$

Em que  $f(\boldsymbol{\theta})$  e  $f(\mathbf{y})$ , são as densidades marginais de  $\boldsymbol{\theta}$  e  $\mathbf{y}$ , respectivamente. Deseja-se estimar  $\boldsymbol{\theta}$  dado a distribuição de  $\mathbf{y}$ , assim, fazendo [4] em [3], tem-se que:

$$f(\boldsymbol{\theta}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})/f(\mathbf{y}) \quad [5]$$

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

$$f(\mathbf{y}) = \int_{R_{\theta}} f(\mathbf{y}, \boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{R_{\theta}} f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta} = E_{\theta} [f(\mathbf{y}|\boldsymbol{\theta})] \quad [6]$$

em que  $E_{\theta}$  representa a média com respeito à distribuição de  $\boldsymbol{\theta}$ . Nota-se, portanto que  $f(\mathbf{y})$  não é uma função de  $\boldsymbol{\theta}$ . Logo [5] pode ser escrita em proporcionalidade (GIANOLA E FERNANDO, 1986) como:

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \quad [7]$$

Pela terminologia Bayesiana,  $f(\boldsymbol{\theta})$  é conhecida como a densidade a *priori* de  $\boldsymbol{\theta}$ , que reflete a incerteza relativa sobre os possíveis valores de  $\boldsymbol{\theta}$  antes das observações serem realizadas, ou seja, antes de se conhecer os valores do vetor  $\mathbf{y}$ . A densidade  $f(\mathbf{y}|\boldsymbol{\theta})$  é a função de verossimilhança que representa a contribuição de  $\mathbf{y}$  para o conhecimento de  $\boldsymbol{\theta}$  e  $f(\boldsymbol{\theta}|\mathbf{y})$  é a densidade a *posteriori* que considera o estado de incerteza sobre  $\boldsymbol{\theta}$  dado um conhecimento prévio.

As inferências Bayesianas são realizadas com base em densidades marginais a *posteriori* dos parâmetros de interesse, obtidas a partir da distribuição conjunta a *posteriori* completa, em que todos os efeitos considerados no modelo devem ser incluídos. Sua obtenção no contexto do melhoramento animal é apresentada a seguir.

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

$$f(b_1|\mathbf{y}), f(b_2|\mathbf{y}), \dots, f(a_1|\mathbf{y}), f(a_2|\mathbf{y}), \dots, f(\sigma_a^2|\mathbf{y}), f(\sigma_e^2|\mathbf{y}) \quad [8]$$

Assim como combinações desses parâmetros, como:

$$f[\sigma_a^2 / (\sigma_a^2 + \sigma_e^2) | \mathbf{y}] = f(h^2 | \mathbf{y}) \quad [9]$$

Estas amostras serão utilizadas para inferências, e para isso, amostras aleatórias da distribuição conjunta a *posteriori* completa são inicialmente acessadas:

$$f(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}) \quad [10]$$

Esta função conduzirá 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 obtê-la, são necessárias a função de verossimilhança e as distribuições a *priori* conforme denotado em [7], uma vez que:

$$f(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2) f(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2) \quad [11]$$

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

$$f(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2) f(\mathbf{b}) f(\mathbf{a}) f(\sigma_a^2) f(\sigma_e^2) \quad [12]$$

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

$$y|\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 \sim N(\mathbf{Xb} + \mathbf{Za}, \mathbf{I} \sigma_e^2) \quad [13]$$

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

$$f(\mathbf{b}) \propto \text{constante}. \quad [14]$$

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

$$\mathbf{a}|\mathbf{A}, \sigma_a^2 \sim N(0, \mathbf{A} \sigma_a^2) \quad [15]$$

Completando 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:

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

em que:

$\sigma_i^2$  = componente de variância  $i$ ;

$\nu_i$  = 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  $i$ .

Desta forma a distribuição conjunta *a posteriori* [12] é representada por:

$$f(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}, \mathbf{v}_i, \mathbf{s}_i^2) \propto (\sigma_e^2)^{-\left(\frac{n+v_e}{2}+1\right)} \exp \left[ -\frac{(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})' (\mathbf{y} - \mathbf{Xb} - \mathbf{Za}) + v_e s_e^2}{2\sigma_e^2} \right] \\ \times (\sigma_a^2)^{-\left(\frac{q+v_a}{2}+1\right)} \exp \left[ -\frac{\mathbf{a}' \mathbf{A}^{-1} \mathbf{a} + v_a s_a^2}{2\sigma_a^2} \right], \quad [17]$$

em que  $n$  e  $q$  são o número total de observações e o número de elementos do vetor  $\mathbf{a}$ , respectivamente.

Normalmente, não é possível obter amostras aleatórias diretamente da distribuição *a posteriori* completa [17] porque se trata de uma função de alta dimensão, sendo o produto de diversas funções multidimensionais. Para reduzir a dimensionalidade e tornar possível a obtenção de amostras aleatórias, são utilizadas técnicas MCMC (Monte Carlo Markov Chain). Dentre elas a mais empregada tem sido o Amostrador de Gibbs cuja técnica é voltada para a 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 a implementação do Amostrador de Gibbs, devem-se derivar todas as distribuições condicionais *a posteriori* a partir da conjunta apresentada em [17], cujo resultado pode ser encontrado em WANG *et al.* (1994) 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 apresentado a seguir:

Define-se a função condicional:

$$f(b_1 | b_2, b_3, \dots, b_i, a_1, a_2, \dots, a_i, \sigma_a^2, \sigma_e^2, \mathbf{y}) \quad [18]$$

em seguida, são fornecidos valores arbitrários para  $b_2, b_3, \dots, b_i, a_1, a_2, \dots, a_i, \sigma_a^2, \sigma_e^2$ . Amostra-se, então, um valor aleatório para  $b_1$  desta distribuição condicional. Com o valor de  $b_1$  aleatoriamente amostrado e

os demais valores arbitrários, tira-se uma amostra aleatória de  $b_2$  da distribuição condicional:

$$f(b_2 | b_1, b_3, \dots, b_j, a_1, a_2, \dots, a_j, \sigma_a^2, \sigma_e^2, \mathbf{y}) \quad [19]$$

Com os valores amostrados para  $b_1$ ,  $b_2$  e os valores arbitrários iniciais, retira-se uma amostra aleatória de  $b_3$ , e procede-se desta forma até que uma amostra de cada parâmetro desconhecido tenha sido tomada de cada distribuição condicional. Após este 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).

### **1.3. Avaliação genética da raça Pardo-suíça.**

O primeiro trabalho de avaliação genética da raça no Brasil descrito na literatura foi realizado por COELHO *et al.* (1993), avaliando 385 lactações de vacas filhas de 10 touros. Neste estudo foram obtidas estimativas de  $0,37 \pm 0,21$  para produção de leite e  $0,21 \pm 0,16$  para produção de gordura.

Paralelamente, SANTUS *et al.* (1993) estimaram parâmetros genéticos em rebanhos italianos de Pardo-suíço utilizando 369.922 registros de primeira lactação de vacas filhas. Os dados foram provenientes da Associazione Nazionale Allevatori Razza Bruna Research Office, até então nada havia sido feito quanto à estimação de componentes de variância sobre esta população. As estimativas de herdabilidade obtidas foram de 0,28 e 0,30 para produção de leite e gordura respectivamente.

RENNÓ *et al.* (2002) estimaram parâmetros genéticos de 11189 lactações de vacas Pardo-suíças de registros provenientes da Associação Brasileira de Criadores de Gado Pardo-suíço. Lactações de animais com registros entre os anos de 1980 e 1999 foram ajustadas por meio de fatores multiplicativos de ajustamento para duas ordenhas, períodos de lactação de 305 dias e produção à idade adulta. As médias e desvios-padrão encontrados para produção de leite e gordura foram respectivamente  $5.879,85 \pm 1.655,79$  kg e  $221,03 \pm 62,52$  kg. As estimativas de herdabilidade foram 0,37 para produção de leite e 0,36 para produção de gordura, com correlação genética de 0,96.

SIROL (2002) avaliando rebanhos nacionais de Pardo-Suíço obteve correlação genética de 0,93. Para verificar a classificação de touros para produção de gordura quando estes são selecionados para leite, o autor fez uso de correlações de Pearson e Spearman. Verificou-se na análise com 10% dos reprodutores que, dos sete considerados como melhores para produção de leite, quatro estavam entre os melhores para produção de gordura.

SIROL *et al.* (2005) avaliando o efeito da interação reprodutor x rebanho sobre a produção de leite e de gordura na raça Pardo-suíça, verificou que a interação não afetou nas estimativas dos componentes de variância, os valores de herdabilidades obtidos com interação foram 0,38 e 0,39 para produção de leite e de gordura respectivamente.

BUENO *et al.* (2005) verificou o efeito das interações reprodutor x rebanho e reprodutor x rebanho-ano sobre os valores genéticos de reprodutores da raça Pardo-Suíça. As estimativas dos componentes de (co)variâncias genética aditiva e residuais praticamente não alteraram quando os modelos foram ajustados para os efeitos de interação. As estimativas de herdabilidade tanto para produção de leite como para de gordura foi 0,40 e correlação genética de 0,94 entre as características exceto quando o modelo considerou o efeito da interação reprodutor x rebanho. Neste caso, para produção de gordura, a herdabilidade foi 0,39 e correlação genética entre as características de 0,95.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, R.D. Variance components. In: USE of mixed model for prediction and for estimation of (co)variance components. Armidale: University of New England, AGBU, 1984. p 77-145.

AZZALINI, A.; VALLE, A.D. The multivariate skewnormal distribution. *Biometrika*, v.83, n.4, p.715-726, 1996.

BLASCO, A. The Bayesian controversy in animal breeding. *J. Anim. Sci.*, v.79, n.8, p.2023-2046, 2001.

BOLDMAN, K.G.; KRIESE L.A.; Van VLECK, L.D.; Van TASSELL, C.P.; KACHMAN, S.D. A Manual for Use of MTDFREML. A set of programs to obtain estimates of variance and covariance [Draft]. Lincoln, U.S. Department of Agriculture, Agriculture Research Service, 1995. 115p.

BROWNIE, C.; BOOS, D.D.; OLIVER, J.H. Modifying the t and ANOVA F tests when treatment is expected to increase variability relative to controls. *Biometrics*, v.46, n.1, p.259-266, 1990.

BUENO, Rachel Santos, TORRES, Robledo de Almeida, RENNO, Francisco Palma *et al.* Efeito da interação reprodutor x rebanho sobre os valores genéticos de reprodutores para produção de leite e gordura na raça Pardo-Suíça. *R. Bras. Zootec.*, jul./ago. 2005, vol.34, no.4, p.1156-1164. ISSN 1516-3598.

CANTET, R.J.C.; FERNANDO R.L.; GIANOLA, D. Bayesian inference about dispersion parameters of univariate mixed models with maternal effects: Theoretical considerations. *Genet. Sel. Evol.*, v.24, p.107-135, 1992.

CASELLA, G.; GEORGE, E.I. Explaining the Gibbs Sampler. *The American Statistician*, v.46, n.3, p.167-174, 1992.

COCHRAN, W.G.; COX, D.F. *Desenho experimentales*. Mexico: Trillas. 1978. 661p.

COELHO, M. M. Neiva, R S; Oliveira, A I G; Silva, A R P; Silva, H C M; Packer, I U; . Fatores de meio e geneticos em caracteristicas produtivas e reprodutivas nas racas holandesa e Pardo-suiça. Iv. Estimativa de parametros geneticos. Lavras, Revista da Sociedade Brasileira de Zootecnia Vicosa, v.22, n.3 , p.445-50, mai./jun. 1993.

FREITAS, A.R. Avaliação de procedimentos na estimação de parâmetros genéticos em bovinos de corte. R. Bras. Zootec., v.29, n.1, p.94-102, 2000.

GELFAND, A.E.; SAHU, S.K.; CARLIN, B.P. Efficient parameterisations for normal linear mixed models. *Biometrika*. v.82, p.479-488, 1995.

GELMAN, A.J.; CARLIN, H.S.; STERN, D.B. Bayesian data analysis. Chapman and Hall, London. 1995.

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.; FOULLEY, J.L.; FERNANDO, R.L. et al. Estimation of heterogeneous variances using empirical bayes methods: Theoretical considerations. J. Dairy Sci., v.75, n.10, p.2805-2823, 1992.

GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theories. J. Anim. Sci., v.63, p.217-244, 1986.

GIANOLA, D.; FOULLEY, J.L. Non linear prediction of latent genetic liability with binary expression: An empirical Bayes approach. In: Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain. v.7, p.293-303, 1982.

GRAZER, H.U.; SMITH, S.P.; TIER, B. A derivate-free approach for estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci., v.64, n.5, p.1362-1370, 1987.

HARTLEY, H.O., RAO, J.N.K. 1967. Maximum likelihood estimation for the mixed analysis of variane model. *Biometrika*, 54(1/2): 93-108.

HASTINGS, W.K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57:97-109.

HENDERSON, C.R. Recent developments in variance and covariance estimation. *J. Anim. Sci.*, v.63, n.1, p.208-216. 1986.

Kennedy, B.W. (1991). C.R. Henderson: the unfinished legacy. *J. Dairy Sci.*, 74:4067-4081.

MEYER, K. Between algorithms: A "Short Cut" restricted maximum likelihood procedure to estimate variance components. *J. Dairy Sci.*, v.69, n.7, p.1904-1916, 1986.

PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block size are unequal. *Biometrika*, v.58, n.3, p.545-554, 1971.

RENNO, Francisco Palma, PEREIRA, José Carlos, ARAUJO, Cláudio Vieira de et al. Aspectos produtivos da raça Pardo-Suíça no Brasil: Fatores de ajustamento, produção de leite e de gordura, e parâmetros genéticos. *R. Bras. Zootec.*, set./out. 2002, vol.31, no.5, p.2043-2054. ISSN 1516-3598.

ROTHSCHILD, M.F.; HENDERSON, C.R.; QUAAS, R.L. Effects of selection on variances and covariances of simulated first and second lactations. *J. Dairy Sci.*, v.62, p.996-1002, 1979.

SANTUS, E. C.; EVERET, R. W.; QUASS, R. L., GALTON, D. M.. Genetic parameters of italian Brown swiss for levels of herds. *Journal of Dairy Science*, 1993, vol. 76, no. 11, p. 3594-3600,

SCHEFFÉ, A. The analysis of variances. [S.l.]: John Wiley. 1959. 477p.

SHOEMAKER, J.S.; PAINTER, I.S.; WEIR, B.S. Bayesian statistics in genetics. *Bayesian Statistical Methods*. v.15, n.9, p.354-358, 1999.

SIROL, M. L. F. G. Efeito da interação reprodutor x rebanho sobre a produção e gordura do leite. Viçosa, MG: Universidade Federal de Viçosa, 2002. 68 p. Dissertação (Mestrado em Zootecnia) – Universidade Federal de Viçosa, 2002.

SIROL, Mirella Leme Franco Geraldini, EUCLYDES, Ricardo Frederico, TORRES, Robledo de Almeida *et al.* Efeito da interação reprodutor x rebanho sobre as produções de leite e de gordura na raça Pardo-Suíça. *R. Bras. Zootec.*, set./out. 2005, vol.34, no.5, p.1573-1580. ISSN 1516-3598.

SORENSEN, D.; GIANOLA, D. (2002). *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Springer, USA, 740p.

SORENSEN, D.A.; WANG, C.S.; JENSEN, J.; GIANOLA, D. Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet. Sel. Evol.* v.26, p.333-360, 1994.

THOMPSON, R. Sire evaluation. *Biometrics*, v.35, p.339-353, 1979.

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 TASSELL, C.P.; VAN VLECK, L.D. A manual for use of MTGSAM. A set of FORTRAN programs to apply Gibbs sampling to animal models for variance components estimation (DRAFT). Lincoln: Department of Agriculture, Agricultural Research Service, 1995, 86p.

WANG, C.S.; GIANOLA, D.; SORENSEN, D.A.; JENSEN, J.; CHRISTENSEN, A.; RUTLEDGE, J.J. Response to selection for litter size in Danish Landrace Pigs: A Bayesian analysis. *Theor. Appl. Genet.*, v.88, p.220-230, 1994.

WOLFINGER, R.D.; KASS, R.E. Nonconjugate Bayesian analysis of variance component models. *Biometrics*. v.56, p.768-774, 2000.

<http://www.cnpqgl.embrapa.br/producao/producao.php>

<http://www.pardo-suico.com.br/>

## CAPÍTULO 1

YAMAKI, Marcos, M. S., Universidade Federal de Viçosa, julho de 2006.  
**Estimação de parâmetros genéticos de produção de leite e de gordura da raça Pardo-suíça, utilizando máxima verossimilhança restrita.**  
Orientador: Robledo de Almeida Torres. Co-orientadores: Paulo Sávio Lopes, Ricardo Frederico Euclides

Dados de primeira lactação de 6.262 vacas distribuídas em 311 rebanhos e filhas de 803 touros com partos entre os anos de 1980 a 2003, foram editados para manter animais com idade ao primeiro parto até 40 meses e com duração de lactação entre 150 e 365 dias, meses de parto das vacas agrupados em quatro estações, grupo contemporâneo formado por rebanho e ano. O objetivo do presente trabalho foi estimar componentes de (co)variância para produção de leite e gordura pelo método da Máxima Verossimilhança Restrita, sob modelo animal, por análises uni e bicaracterística. Foi testado o efeito da inclusão ou não da duração da lactação nos modelos utilizados para as duas características. As estimativas foram obtidas utilizando-se o programa MTDFREML (BOLDMAN *et al.* 1995). Foram testados modelos unicaracterística com diferentes ordens para as covariáveis e considerados grupo contemporâneo e estação como efeitos fixos. Os melhores ajustes obtidos nas análises unicaracterística foram utilizados na análise bicaracterística. A duração da lactação reduziu a estimativa da variância aditiva quando era utilizada no modelo reduzia a estimativa da variância aditiva, sugerindo que os animais estariam sendo corrigidos para uma mesma base quanto à capacidade de imprimir duração da lactação mais longa ou mais curta à progênie sendo, portanto, não recomendado o ajuste para esta covariável. Já a idade da vaca ao parto, influenciou linearmente a produção de leite e gordura. As herdabilidades estimadas foram 0,26 e 0,25 para produção de leite e gordura, respectivamente, com correlação genética de 0,95. A alta correlação entre a produção de leite e gordura obtida sugere que parte dos genes que

atuam na produção de leite também responde pela produção de gordura, de tal forma que a seleção para a produção de leite resulta, indiretamente, em aumentos na produção de gordura.

## ABSTRACT

YAMAKI, Marcos, M. S., Universidade Federal de Viçosa, July 2006.  
**Estimation of genetic parameters of milk and fat yield of Brown-Swiss cows using frequentist methodology.** Adviser: Robledo de Almeida Torres. Co-advisers: Paulo Sávio Lopes and Ricardo Frederico Euclides

First lactation data of 6.262 Brown-Swiss cows from 311 herds, daughters of 803 sires with calving between 1980 and 2003 were used to estimate genetic parameters for milk and fat production traits. The components of variance were estimated by restricted maximum likelihood (REML) methods, using animal model with uni and two-traits analysis. Data was edited to maintain animals with age of calving between 150 and 365 days, months of calving were grouped in 4 seasons, contemporary group was formed by herd and year. The estimation by REML was obtained with the software MTDFREML (BOLDMAN *et al.* 1995) testing unitrait models with different effects to covariables and considering contemporary group and season as fixed effect. The best fitting obtained on unitrait analysis were used on two-trait analysis. The estimative of additive variance was reduced when lactation length was included on the model suggesting that the animals were been fitted to the same base on the capacity of transmit a longer or shorter lactation length to the progeny. Therefore, fitting to this covariable is not recommended. On the other side, the age of calving has linearly influenced milk and fat production. The heritability estimates were 0,26 and 0,25 to milk and fat yield respectively with genetic correlation of 0,95. the high correlation among these traits suggests that part of genes that acts on milk yield also respond to fat yield, in such way that selection for milk yield results, indirectly, in increase on fat yield. The estimation by Bayesian inference was made on software MTGSAM – (VAN TASSELL E VAN VLECK, 1995). Chain lengths were tested to obtain the marginal posterior densities of unitrait analysis, the best option of chain length, burn-in and sampling interval was used on two-trait analysis. The burn-in periods were tested with the software GIBANAL (VAN KAAM, 1998) witch analysis inform a sampling interval for each burn-in tested, the criteria for choosing the sampling interval was made with the serial correlation resulting by the bur-in and the sampling process. The heritability estimates were  $0,33 \pm 0,05$  for both traits with genetic correlation of

0,95. Similar results were obtained on studies using the same methodology on first lactation records. The stationary phase adequately reached with a 500.000 chain length and 30.000 burn-in interactions.

## 1. INTRODUÇÃO

O conhecimento do potencial genético dos animais destinados à produção de leite e gordura, bem como os fatores que influenciam a expressão destas duas características são de suma importância na tomada de decisão das estratégias de seleção em um programa de melhoramento. A correlação genética existente entre estas características também constitui uma importante fonte de informação quando o objetivo do sistema de produção não se destina somente à produção de leite, mas também para a produção de seus derivados como o queijo e a manteiga.

Diante das necessidades particulares dos diversos sistemas de produção leiteira no Brasil, a indicação da raça ou cruzamento de raças para tal deve ser feita de forma criteriosa, sob pena de se obter índices de produtividade baixos caso a recomendação não seja adequada ou ainda, se a avaliação dos animais for feita de forma errônea. Assim, deve-se procurar levar em consideração os fatores que têm grande influência nas características em questão.

Como exemplo de fatores que influenciam na produção de leite e gordura, pode-se citar: ano e estação de parto, rebanho, período de lactação e idade da vaca ao parto.

A influência do efeito de ano sobre a produção de leite e gordura reflete de diversas maneiras que podem ser tanto de ordem ambiental (manejo, alimentação) quanto econômica (rentabilidade da atividade leiteira). ALMEIDA *et al.* (1999) verificaram efeitos de ano de parto sobre a produção de leite e gordura em rebanhos da raça Pardo-suíça no estado do Paraná.

A estação de parto influencia a produção de leite e gordura principalmente pela qualidade e quantidade de alimento, principalmente de forragem, ofertada aos animais durante as quatro estações do ano. As

condições climáticas também são consideradas neste fator influenciando a produção direta ou indiretamente pela temperatura, umidade relativa do ar, fotoperíodo e outros fatores que alteram no decorrer das estações. ALMEIDA *et al.* (1999) e BAJALUK *et al.* (1999) verificaram o efeito da época de parto sobre as características produtivas estudadas nos rebanhos da raça Pardo-suíça e holandesa respectivamente.

O efeito de rebanho pode ser explicado, principalmente, devido a variações no manejo, localização geográfica, composição genética da população e de critérios de seleção. ALMEIDA *et al.* (1999) e BAJALUK *et al.* (1999) também verificaram efeito de rebanho sobre as produções de leite e gordura.

Segundo RENNÓ (2001), a idade ao primeiro parto e a produção de leite na primeira lactação, são de extrema importância dentro dos sistemas de produção de leite. A idade ao primeiro parto, além de ser considerada um índice reprodutivo, também deve ser considerada um parâmetro para avaliar as atividades relacionadas com o manejo das categorias jovens do rebanho. Assim o estudo da idade ao primeiro parto aliado à produção na primeira lactação é de grande valor para avaliar e estabelecer o manejo do rebanho.

A duração da lactação influi diretamente na produção de leite na lactação, quanto maior a duração da lactação, maior a produção de leite na lactação. Entretanto períodos muito longos dificultam o manejo do rebanho, o período considerado ideal está em torno dos 305 dias, levando-se em conta um período de descanso de 60 dias o que proporciona um parto por ano. O que é comumente feito na avaliação da produção de leite é ajustar a lactação para 305 dias, no entanto esta prática pode prejudicar vacas boas produtoras de leite e favorecer vacas ruins, principalmente quando se está avaliando animais de primeira lactação, pois não há como avaliar repetibilidade da duração da lactação.

O objetivo deste estudo foi estimar parâmetros genéticos produção de leite e gordura em registros de primeira lactação de animais da raça Pardo-suíça.

## 2. MATERIAIS E MÉTODOS

Foram utilizados dados provenientes do Controle Leiteiro da Associação Brasileira dos Criadores de Gado Pardo-Suíço (ABCGPS), fornecidos por intermédio do convênio mantido entre o Departamento de Zootecnia da Universidade Federal de Viçosa e a ABCGPS. O banco de dados inicial era constituído de informações de primeira lactação de 6.262 vacas distribuídas em 311 rebanhos e filhas de 803 touros com partos entre os anos de 1980 a 2003.

Os dados foram editados para manter animais com idade ao primeiro parto até 40 meses e com duração de lactação entre 150 e 365 dias.

Os meses de parto das vacas foram agrupados em quatro estações; estação 1, correspondente aos meses de janeiro a março; estação 2, de abril a junho; estação 3, de julho a setembro e estação 4, de outubro a dezembro. As estações 1 e 4 correspondem aos meses de maior incidência de chuvas, maiores temperaturas e umidade relativa do ar, uma vez que a maior parte dos dados corresponde a rebanhos localizados na região Sul e Sudeste.

Os animais foram reunidos em grupos contemporâneos formados com classes de rebanho e ano de nascimento, sendo considerados apenas aqueles que possuíam pelo menos três animais. O efeito de estação foi considerado como efeito fixo separadamente no modelo por formar muitos grupos contemporâneos com apenas um animal quando eram criadas classes deste efeito juntamente com os efeitos de rebanho e ano

Após as restrições, o banco de dados constituiu de 2758 registros de primeira lactação de animais de vacas, filhas de 151 pais, distribuídas em 339 grupos contemporâneos. A distribuição dos animais nas estações pré-definidas é apresentada na Tabela 1:

Tabela 1 – Distribuição das vacas por estação de parto.

<b>Estação<sup>1</sup></b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>total</b>
<b>Número de animais</b>	609	778	763	608	2.758

Estação 1: período compreendido de janeiro à março; estação 2; de abril a junho; estação 3: de julho à setembro e estação 4: de outubro à dezembro

Os componentes de (co)variância foram estimados pelo método da máxima verossimilhança restrita, utilizando-se o programa MTDFREML (“Multiple Trait Derivate Free Restricted Maximum Likelihood”), descrito por BOLDMAN *et al.* (1995), utilizando-se o modelo animal. O programa utiliza o algoritmo simplex para localizar o mínimo de  $-2 \log_e L$ , em que L é a função de verossimilhança, e os componentes de (co)variância que minimizam a função  $-2 \log_e L$  são estimativas de verossimilhança.

Como critério de convergência, foi utilizado a variância dos valores do simplex ( $-2 \log_e L$ ) iguais a  $10^{-9}$ .

As análises foram realizadas utilizando-se modelos uni e bicaracterística para produção e gordura do leite.

O modelo utilizado nas análises é representado por:

$$\tilde{y} = X \tilde{b} + Z \tilde{a} + \tilde{e}; \text{ em que:}$$

$\tilde{y}$  = vetor n x 1 de observações;

$X$  = matriz n x f de incidência de efeitos fixos de grupo contemporâneo, estação de parto e covariáveis;

$\tilde{b}$  = vetor f x 1 de efeitos fixos;

$Z$  = matriz n x N de incidência de efeitos genéticos aditivos diretos;

$\tilde{a}$  = vetor N x 1 de efeitos genéticos aditivos diretos;

$\tilde{e}$  = vetor de efeito residual de mesma dimensão de  $\tilde{y}$ .

Sendo

N = número de indivíduos

n = número total de observações; e

f = número de classes de efeitos fixos

As pressuposições para os efeitos aleatórios são:

$$\begin{bmatrix} \tilde{a} \\ \tilde{e} \end{bmatrix} \sim N \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} G & \phi \\ \phi & R \end{bmatrix} \right\}, \text{ em que}$$

G é a matriz de (co)variâncias genéticas aditivas diretas,  $G = A \sigma_a^2$ ; A é matriz de numeradores dos coeficientes de parentesco de Wright entre os

indivíduos;  $R$  é a matriz de (co)variâncias residuais;  $R = I\sigma_e^2$  e  $\phi$  é a matriz de zeros.

Foram testados vários modelos nas análises unicaracterística de produção de leite e gordura, aquele que melhor se ajustou foi utilizado na análise bicaracterística.

O modelo inicial utilizado para as análises das duas características continha efeito fixo de grupo contemporâneo e estação, covariáveis idade da vaca ao parto e duração da lactação com efeitos linear e quadrático. A partir daí, foram comparados modelos com ou sem a inclusão da duração da lactação com diferentes coeficientes de regressão. O motivo de se testar a duração da lactação no modelo se deve a relatos na literatura de que, em bovinos leiteiros, o ajuste para duração da lactação diminui a variabilidade genética para produção de leite (MELLO *et al.*, 1994).

Na Tabela 2 estão descritos os modelos utilizados nas análises.

Tabela 2 – Disposição dos efeitos e dos coeficientes de regressão linear (I) e quadrático (II) utilizados nos modelos das análises.

MODELOS	GC <sup>1</sup>	ESTAÇÃO	DL <sup>2</sup>		IP <sup>3</sup>	
			I	II	I	II
1	X	X		X		X
2	X	X		X	X	
3	X	X	X		X	
4	X	X				X
5	X	X			X	
6	X				X	

1 – Grupo contemporâneo; 2 – Duração da lactação; 3 – Idade da vaca ao parto

O critério de escolha do modelo mais adequado foi feito segundo a contribuição ou não do efeito para a variância genética aditiva.

### 3. RESULTADOS E DISCUSSÃO

A tabela a seguir apresenta uma descrição dos dados de produção de leite e gordura bem como da duração da lactação e idade da vaca ao parto.

Tabela 3 – Número de observações (N), desvios-padrão (DP), mínimos e máximos da produção de leite e de gordura, duração da lactação (DL) e idade da vaca ao parto (IP).

	<b>N</b>	<b>Médias</b>	<b>DP</b>	<b>mínimo</b>	<b>máximo</b>
<b>PL</b> <sup>1</sup>	2.758	5.392,32	1.852,13	992	13.479
<b>PG</b> <sup>2</sup>	2.758	203,42	71,02	34	493,80
<b>DL</b> <sup>3</sup>	2.758	310,23	59,82	150	365
<b>IP</b> <sup>4</sup>	2.758	29,61	4,52	19	40

1 – Produção de leite em kg; 2 – Produção de gordura em kg; 3 – Duração da lactação em dias; 4 – Idade da vaca ao parto em meses

SANTUS *et al.* (1993), avaliando rebanhos italianos de Pardo-suíço obtiveram resultados similares. Apesar de dispor de, relativamente, poucas observações, o banco de dados em estudo apresentou valores médios muito próximos aos encontrados na literatura em estudos com banco de dados maiores.

RENNÓ *et al.* (2002) estudando aspectos produtivos da raça Pardo-suíça estimaram médias para produção de leite e gordura superiores aos encontrados neste trabalho devido aos fatores de ajuste e à restrição imposta ao banco de dados, da qual foram excluídos lactações inferiores à 1.000kg. Isto também colaborou para melhorar a consistência dos dados que pode ser observado pelos menores desvios quando comparados aos encontrados neste estudo.

RENNÓ (2001) obteve valor médio da idade da vaca ao parto, similar aos neste estudo de  $28,70 \pm 2,86$  kg, avaliando animais da raça entre 1985 a 1996, a média bem como seu desvio apresentam valores menores em função

das restrições impostas pelo autor considerando vacas com idades entre 22 a 35 meses enquanto que este trabalho adotou a idade até 40 meses.

As análises unicaracterística foram iniciadas a partir de um modelo considerando duração da lactação e idade da vaca ao parto como covariáveis e o grupo contemporâneo e a estação como efeitos fixos. A partir daí, os ajustes mais adequados foram utilizados na análise bicaracterística bem como as estimativas de herdabilidade utilizados como valores iniciais.

Os componentes de (co)variâncias e herdabilidades das análises unicaracterística para produção de leite e gordura são apresentados nas Tabelas 4, 5 e 6.

Tabela 4 – Estimativas da variância genética aditiva ( $\sigma_A$ ), residual ( $\sigma_E$ ) e fenotípica ( $\sigma_P$ ) e herdabilidade ( $h^2$ ) da análise unicaracterística de diferentes modelagens para produção de leite.

	<b>Modelos</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
$\sigma_A$	233.104,25	233.217,26	233.648,96	504.012,23	508.410,60	504.890,36
$\sigma_E$	777.795,58	778.758,07	776.507,64	1.437.288,31	1.433.232,80	1.439.947,44
$\sigma_P$	1.010.899,83	1.011.975,33	1.010.156,60	1.941.300,54	1.941.643,40	1.944.837,80
$h^2$	0,23	0,23	0,23	0,26	0,26	0,26

- 1 – Coeficiente linear e quadrático para as covariáveis duração da lactação e idade da vaca ao parto.  
 2 – Coeficiente linear e quadrático para duração da lactação e linear para idade da vaca ao parto.  
 3 – Coeficiente linear para as covariáveis duração da lactação e idade da vaca ao parto.  
 4 – Modelo com idade da vaca ao parto com coeficiente linear e quadrático e desconsiderando duração da lactação.  
 5 – Modelo com idade da vaca ao parto com coeficiente linear e desconsiderando duração da lactação.  
 6 – Modelo com idade da vaca ao parto com coeficiente linear e desconsiderando duração da lactação e o efeito fixo de estação.

Tabela 5 - Estimativas da variância genética ( $\sigma_A$ ), residual ( $\sigma_E$ ) e fenotípica ( $\sigma_P$ ) e herdabilidade ( $h^2$ ) da análise unicaracterística de diferentes modelagens para produção de gordura.

	Modelos					
	1	2	3	4	5	6
$\sigma_A$	373,553	375,570	376,383	718,677	736,309	721,710
$\sigma_E$	1.126,770	1.127,354	1.126,544	2.148,141	2.136,951	2148,655
$\sigma_P$	1.500,325	1.502,923	1.502,927	2.866,817	2.873,260	2870,365
$h^2$	0,25	0,25	0,25	0,25	0,26	0,25

- 1 – Coeficiente linear e quadrático para as covariáveis duração da lactação e idade da vaca ao parto.
- 2 – Coeficiente linear e quadrático para duração da lactação e linear para idade da vaca ao parto.
- 3 – Coeficiente linear para as covariáveis duração da lactação e idade da vaca ao parto.
- 4 – Modelo com idade da vaca ao parto com coeficiente linear e quadrático e desconsiderando duração da lactação.
- 5 – Modelo com idade da vaca ao parto com coeficiente linear e desconsiderando duração da lactação.
- 6 – Modelo com idade da vaca ao parto com coeficiente linear e desconsiderando duração da lactação e o efeito fixo de estação.

A duração da lactação contribuiu negativamente para as estimativas da variância genética aditiva na modelagem tanto para produção de leite quanto para produção de gordura. Isto ocorre porque quando se faz o ajuste para a duração da lactação todos os touros são corrigidos para uma mesma base, ou seja, é retirada a diferença genética entre os touros quanto à capacidade que possuem de transmitir para suas filhas durações da lactação mais curtas ou mais longas, conseqüentemente reduzindo a herdabilidade.

Em ambas as características observa-se que os valores das estimativas das variâncias aditivas praticamente reduzem o valor à metade quando a duração da lactação foi incluída nos modelos. Em conseqüência, observa-se a redução da herdabilidade nos modelos em que foi considerada a duração da lactação.

Segundo MELLO *et al.* (1994), o ajuste para duração da lactação em bovinos leiteiros diminui a variabilidade genética para produção de leite pela alta correlação entre elas. O mesmo foi observado neste estudo, em que todos

os modelos utilizando período de lactação como covariável reduziram drasticamente a estimativa da variância aditiva em relação àqueles que consideraram apenas a idade da vaca ao parto.

Entre os modelos que utilizaram apenas a idade da vaca ao parto como covariável, aquele que utilizou efeito linear apresentou as maiores estimativas da variância genética aditiva para ambas as características. Isto pode ser justificado pelo uso de apenas a informação da primeira lactação das vacas, pois as vacas desta raça atingem o seu potencial produtivo em lactações subseqüentes em torno dos 72 e 69 meses para produção de leite e gordura como observado por RENNÓ *et al.* (2002).

ALMEIDA *et al.* (1995) verificaram efeito linear da idade ao primeiro parto sobre a produção de leite na primeira lactação em rebanhos de vacas Holandesas, de forma que os animais que pariram em idades tardias apresentaram as maiores produções.

Adicionalmente foi testada a inclusão ou não do efeito fixo de estação de parto para as duas características. Para a produção de gordura houve ligeira queda da estimativa da herdabilidade e não variou para a produção de leite. Em ambas as características observaram-se aumento da variância residual quando o efeito de estação não era incluído, comprovando a importância da sua inclusão nos modelos.

Assim optou-se pelo modelo 5, sem a inclusão da duração da lactação com coeficiente de regressão linear para a idade da vaca ao parto, efeitos fixos de grupo contemporâneo e estação para a análise bicaracterística.

Os resultados da análise bicaracterística são apresentados na Tabela 6.

Tabela 6 – Herdabilidade ( $h^2$ ) (co)variância genética, residual e fenotípica e correlação genética da análise bicaracterística para produção de leite e gordura.

	PL	PG
$h^2$	0,26	0,25
$\sigma_A^1$	503602,981	727,212
$\sigma_E^2$	1437123,938	2143,057
$\sigma_P^3$	1940726,920	2870,269
$COV_A^4$	18134,644	
$COV_E^5$	52078,971	
$COV_P^6$	70213,615	
$r_G^7$	0,95	

1 – Variância genética aditiva; 2 – Variância residual; 3 – Variância fenotípica; 4 – Covariância genética aditiva; 5 - Covariância residual; 6 - Covariância fenotípica; 7 – Correlação genética.

As estimativas de herdabilidade obtidas foram de moderada magnitude, com valores menores aos obtidos por SANTUS *et al.* (1993), que avaliaram parâmetros genéticos de rebanhos italianos de Pardo-suíço para níveis de produção com estimativas de herdabilidades de 0,28 e 0,29 para produção de leite e gordura respectivamente.

Apesar de apresentarem herdabilidades próximas, a correlação foi de alta magnitude e superior a encontrada por SANTUS *et al.* (1993), 0,89. No caso da produção de leite e gordura a correlação é de extrema importância quando o objetivo do programa de melhoramento visa não apenas a produção de leite, mas também a produção de gordura. Neste estudo a correlação genética estimada foi de 0,95.

Segundo SIROL *et al.* (2005) ao avaliar de rebanhos nacionais de Pardo-Suíço para a produção de leite e gordura obteve correlação genética de 0,93. As estimativas de herdabilidades obtidas por COELHO *et al.* (1993) ao estudar lactações vacas Pardo-suíço foram  $0,37 \pm 0,21$  e  $0,21 \pm 0,16$  para produção de leite e gordura, respectivamente. Apesar de apresentar um valor próximo ao encontrado para produção de gordura observa-se grande variação

do resultado possivelmente devido à pequena quantidade de observações utilizadas pelos autores.

RENNÓ *et al.* (2002) ao avaliar o desempenho produtivo de vacas Pardo-suíço, obteve correlação genética de 0,96, similar a este estudo, e as respectivas herdabilidades para as características produção de leite e gordura: 0,37 e 0,36. Os valores das estimativas de herdabilidade foram maiores que os obtidos neste estudo provavelmente em função dos fatores de ajustamento utilizados pelos autores como lactação ajustada para 305 dias e à idade adulta além de restrições feitas no banco de dados. Desta forma, as produções serão subestimadas nos animais com maior persistência de lactação e superestimadas naqueles com menor persistência de lactação (JAMROZIK & SCHAEFFER, 1997).

SIROL *et al.* (2005) ao avaliar o efeito da interação reprodutor x rebanho sobre a produção de leite e gordura na raça Pardo-suíça obteve maiores valores de estimativas de herdabilidade, de 0,38 e 0,39 para produção de leite e gordura respectivamente.

De maneira geral as estimativas para herdabilidade e obtidas neste estudo, tanto para produção de leite quanto gordura, foram menores com relação aos estudos feitos na raça anteriormente, no entanto a estimativa da correlação, considerada alta, foi bem próxima dos valores encontrados na literatura.

#### **4. CONCLUSÃO**

A idade da vaca ao parto influenciou linearmente a produção de leite e gordura, enquanto que a inclusão duração da lactação como covariável contribuiu negativamente para as estimativas de herdabilidade.

Os menores valores das estimativas de herdabilidades comparados à literatura devem-se, possivelmente à ausência de ajustes dos dados de produção.

A alta correlação genética entre a produção de leite e gordura obtida sugere que parte dos genes que atuam na produção de leite também responde pela produção de gordura, de tal forma que a seleção para a produção de leite resulta, indiretamente, em aumentos na produção de gordura.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, R., RIBAS, N. P., MARCONDES, E. A. Estudo de alguns fatores do meio ambiente sobre a produção de leite, gordura e proteína em vacas da raça Pardo-suíça no estado do Paraná. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 36, 1999, Porto Alegre. Anais SBZ, Porto Alegre-RS, 1999, p159.

ALMEIDA, R., RIBAS, N. P., MONARDES, H. G. Estudo das características produtivas em rebanhos Holandeses em primeira cria na Região Batavo, Paraná. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 32, 1995, Brasília. Anais SBZ, Brasília-DF, 1995, p.162.

BAJALUK, S. A. B., RIBAS, N. P., MONARDES, H. G. ET AL. Efeito de fatores ambientais sobre a produção de leite, percentagem de gordura e percentagem de proteína em vacas da raça holandesa no Estado do Paraná. In. REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, Porto Alegre. Anais SBZ, Porto Alegre-RS, 1999, p158.

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.

COELHO, M. M. Neiva, R S; Oliveira, A I G; Silva, A R P; Silva, H C M; Packer, I U; . Fatores de meio e geneticos em caracteristicas produtivas e reprodutivas nas racas holandesa e Pardo-suíça. Iv. Estimativa de parametros geneticos. Lavras, Revista da Sociedade Brasileira de Zootecnia Vicosa, v.22, n.3 , p.445-50, mai./jun. 1993.

JAMROZIK, J.; SCHAEFFER, L.R. Estimates of genetic parameters for a test day model with random regression for yield traits. Journal of Dairy Science, v.80, p.762-770, 1997.

MELLO,A.A.; PENNA, V.M.; MADALENA, F.E. et al. Efeito da eliminação de lactações curtas e do ajuste pela duração da lactação na herdabilidade da produção de leite em um rebanho Gir. *Archivos Latinoamericanos Producción Animal*, v.2, p.117-123, 1994.

RENNÓ, F. P. Aspectos produtivos da raça Pardo-Suíça no Brasil. Viçosa, MG: Universidade Federal de Viçosa, 2001. 100 p. Dissertação (Mestrado em Zootecnia) – Universidade Federal de Viçosa, 2001.

RENNÓ, F. P., PEREIRA, J. C., ARAUJO, C. V. de et al. Aspectos produtivos da raça Pardo-Suíça no Brasil: Fatores de ajustamento, produção de leite e de gordura, e parâmetros genéticos. *R. Bras. Zootec.*, set./out. 2002, vol.31, no.5, p.2043-2054. ISSN 1516-3598.

SIROL, M. L. F. G., EUCLYDES, R. F., TORRES, R. A. *et al.* Efeito da interação reprodutor x rebanho sobre as produções de leite e de gordura na raça Pardo-Suíça. *R. Bras. Zootec.*, set./out. 2005, vol.34, no.5, p.1573-1580. ISSN 1516-3598.

SANTUS, E. C., EVERET, R. W., QUASS, R. L., GALTON, D. M.. Genetic parameters of italian Brown swiss for levels of herds. *Journal of Dairy Science*, 1993, vol. 76, no. 11, p. 3594-3600.

## CAPÍTULO 2

YAMAKI, Marcos, M. S., Universidade Federal de Viçosa, julho de 2006.  
**Estimação de parâmetros genéticos de produção de leite e de gordura da raça Pardo-suíça, utilizando inferência Bayesiana.** Orientador: Robledo de Almeida Torres. Co-orientadores: Paulo Sávio Lopes, Ricardo Frederico Euclides

Dados de primeira lactação de 6.262 vacas distribuídas em 311 rebanhos, filhas de 803 touros com partos entre os anos de 1980 e 2003 foram utilizados para estimar parâmetros genéticos para as características de produção de leite e gordura com informações de primeira lactação em animais da raça Pardo-Suíço por inferência Bayesiana, e ainda, determinar tamanhos de cadeia, burn-in e amostragem adequados para análises uni e bicaracterística. Os dados foram editados para manter animais com idade ao primeiro parto até 40 meses e com duração da lactação entre 150 e 365 dias, meses de parto das vacas agrupadas em quatro estações e grupo contemporâneo formado por rebanho e ano. Foram testados diversos tamanhos de cadeia para a obtenção das densidades marginais a *posteriori* das análises unicaracterística, a melhor proposta para o tamanho de cadeia, *burn-in* e amostragem foi utilizada para a análise bicaracterística. Os períodos de *burn-in* foram testados pelo programa GIBANAL (VAN KAAM, 1998) cujas análises fornecem um intervalo de amostragem para cada *burn-in* testado, o critério de escolha do intervalo de amostragem foi feito de acordo com a correlação serial, resultante do *burn-in* e do processo de amostragem. As estimativas de herdabilidade foram  $0,33 \pm 0,05$  para ambas as características com correlação de 0,95. Resultados similares foram obtidos em estudos utilizando a mesma metodologia em informações de primeira lactação. A fase estacionária foi adequadamente obtida com uma cadeia de 500.000 iterações e descarte inicial de 30.000 iterações.

## Abstract

YAMAKI, Marcos, M. S., Universidade Federal de Viçosa, July 2006.  
**Estimation of genetic parameters of milk and fat yield of Brown-Swiss cows using bayesian methodologys.** Adviser: Robledo de Almeida Torres. Co-adviser: Paulo Sávio Lopes and Ricardo Frederico Euclides

First lactation data of 6.262 Brown-Swiss cows from 311 herds, daughters of 803 sires with calving between 1980 and 2003 were used to estimate genetic parameters for milk and fat production traits by Bayesian inference using animal model with uni and two-traits analysis. Data was edited to maintain animals with age of calving between 150 and 365 days, months of calving were grouped in 4 seasons, contemporary group was formed by herd and year. The components of variance were estimated on software MTGSAM – (VAN TASSELL E VAN VLECK, 1995). Chain lengths were tested to obtain the marginal posterior densities of unitrait analisys, the best option of chain length, burn-in and sampling interval was used on two-trait analisys. The burn-in periods were tested with the software GIBANAL (VAN KAAM, 1998) witch analysis inform a sampling interval for each burn-in tested, the criteria for choosing the sampling interval was made with the serial correlation resulting by burn-in and sampling process. The heritability estimates were  $0,33 \pm 0,05$  for both traits with genetic correlation of 0,95. Similar results were obtained on studies using the same methodology on first lactation records. The stationary phase adequately reached with a 500.000 chain length and 30.000 burn-in interactions.

## 1. INTRODUÇÃO

O progresso genético nos sistemas de produção animal exigem avaliações genéticas acuradas dos animais, seguidas da seleção dos melhores indivíduos que serão pais da próxima geração. Desta forma, há constante interesse por metodologias ou estratégias que produzam estimativas acuradas dos componentes de variância para as diversas situações encontradas na prática.

Dentre os métodos de estimação de componentes de variância, destacam-se a Máxima Verossimilhança Restrita (REML) e a inferência Bayesiana. Os métodos Bayesianos foram introduzidos no melhoramento animal por GIANOLA & FOULLEY *et al.* (1982) e posteriormente por GIANOLA & FERNANDO (1986). As técnicas não foram inicialmente implementadas devido a problemas de resolução de múltiplas integrais; no entanto, os métodos MCMC (Monte Carlo Markov Chain) têm dado solução a muitos problemas que não foram resolvidos anteriormente devido à resolução de tais integrais.

Os métodos MCMC são um conjunto de processos iterativos para a geração aproximada de amostras de distribuições multivariadas. Neles, o conhecimento da variável aleatória gerada só é relevante para a geração da variável seguinte, sendo, portanto, irrelevante na predição das variáveis futuras. Um método MCMC bastante utilizado é o Algoritmo de Metroplis-Hastings (HASTINGS, 1970), sendo o Amostrador de Gibbs (GEMAN &

GEMAN, 1984) um caso especial deste algoritmo que vem sendo empregado por diversos autores na avaliação genética animal.

A maior dificuldade encontrada atualmente é a detecção da convergência da(s) cadeia(s) de Gibbs. Assim, os estudos sobre inferência Bayesiana têm se voltado para a verificação da convergência de métodos de integração numérica. Uma recomendação comum é descartar a parte inicial da cadeia, a qual é denominada *burn-in* e basear os diagnósticos nos elementos restantes da(s) cadeia(s). Teoricamente uma cadeia de Markov converge para sua distribuição de equilíbrio. Entretanto, na prática, é difícil certificar a convergência de uma cadeia, gerada por um Amostrador de Gibbs (WANG *et al.*, 1994).

O objetivo deste estudo foi estimar de componentes de variância para as características de produção de leite e gordura com informações de primeira lactação em animais da raça Pardo-Suíço por inferência Bayesiana, e ainda, determinar tamanhos de cadeia, *burn-in* e amostragem adequados para análises uni e bicaracterística.

## **2. MATERIAIS E MÉTODOS**

### **2.1. DESCRIÇÃO DOS DADOS**

Foram utilizados dados provenientes do Controle Leiteiro da Associação Brasileira dos Criadores de Gado Pardo-Suíço (ABCGPS), fornecidos por intermédio do convênio mantido entre o Departamento de Zootecnia da Universidade Federal de Viçosa e a ABCGPS. O banco de dados inicial era constituído de informações de primeira lactação de 6.262 vacas distribuídas em 311 rebanhos e filhas de 803 touros com partos entre os anos de 1980 a 2003.

Os dados foram editados para manter animais com idade ao primeiro parto até 40 meses e com duração de lactação entre 150 e 365 dias.

Os meses de parto das vacas foram agrupados em quatro estações; estação 1, correspondente aos meses de janeiro a março; estação 2, de abril a junho; estação 3, de julho a setembro e estação 4, de outubro a dezembro. As estações 1 e 4 correspondem aos meses de maior incidência de chuvas, maiores temperaturas e umidade relativa do ar, uma vez que a maior parte dos dados corresponde a rebanhos localizados na região Sul e Sudeste.

Os animais foram reunidos em grupos contemporâneos formados com classes de rebanho e ano, sendo considerados apenas aqueles que possuíam pelo menos três animais. O efeito de estação foi considerado como um efeito fixo separadamente no modelo por formar muitos grupos contemporâneos com apenas um animal quando eram criadas classes deste efeito juntamente com os efeitos de rebanho e ano

Após as restrições, o banco de dados constituiu de 2.758 registros de primeira lactação de animais de vacas, filhas de 151 pais, distribuídas em 339 grupos contemporâneos. A distribuição dos animais nas estações pré-definidas é apresentada na Tabela 1:

Tabela 1 – Distribuição das vacas por estação de parto.

Estação <sup>1</sup>	1	2	3	4	total
Número de animais	609	778	763	608	2.758

Estação 1: período compreendido de janeiro à março; estação 2; de abril a junho; estação 3: de julho à setembro e estação 4: de outubro à dezembro

## 2.2. MODELOS

As análises foram realizadas utilizando-se modelos uni e bicaracterística para produção de leite e de gordura.

O modelo utilizado nas análises é representado por:

$$\tilde{y} = X \tilde{b} + Z \tilde{a} + \tilde{e} \quad [1]$$

em que:

$\tilde{y}$  = vetor n x 1 de observações;

$X$  = matriz n x f de incidência de efeitos fixos de grupo contemporâneo, estação de parto e covariáveis;

$\tilde{b}$  = vetor f x 1 de efeitos fixos;

$Z$  = matriz n x N de incidência de efeitos genéticos aditivos diretos;

$\tilde{a}$  = vetor N x 1 de efeitos genéticos aditivos diretos;

$\tilde{e}$  = vetor de efeito residual de mesma dimensão de  $\tilde{y}$ .

Sendo

N = número de indivíduos

n = número total de observações; e

f = número de classes de efeitos fixos

As pressuposições para os efeitos aleatórios são:

$$\begin{bmatrix} y \\ \tilde{a} \\ \tilde{e} \end{bmatrix} \sim N \left\{ \begin{bmatrix} Xb \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} ZGZ'+R & ZG & R \\ GZ' & G & \phi \\ R & \phi & R \end{bmatrix} \right\} \quad [2]$$

em que:

G é a matriz de (co)variâncias genéticas aditivas diretas,  $G = A\sigma_a^2$ ; A é matriz de numeradores dos coeficientes de parentesco de Wright entre os indivíduos; R é a matriz de (co)variâncias residuais;  $R = I\sigma_e^2$  e  $\phi$  é a matriz de zeros.

Para ambas as características, a covariável idade da vaca ao parto foi considerada com coeficiente de regressão linear, tanto nas análises uni quanto bicaracterística.

### 2.3. ANÁLISES

Os componentes de variância foram estimados por inferência Bayesiana utilizando-se o programa MTGSAM – *Multiple Trait Gibbs Sampling in Animal Models* (VAN TASSELL & VAN VLECK, 1995). O MTGSAM é composto por um conjunto de programas escritos em FORTRAN, desenvolvido para estimação de componentes de variância via Amostrador de Gibbs.

Para os componentes de variância foi utilizada uma informação a *priori* do tipo *flat*, ou seja, considerou-se que há um vago conhecimento a *priori* sobre a distribuição dos mesmos. Os valores iniciais da variância genética aditiva e residual utilizados foram obtidos em análises anteriores (capítulo 1) via máxima verossimilhança restrita pelo programa MTDFREML descrito por BOLDMAN *et al.* (1995).

Para a obtenção das densidades marginais a *posteriori* das análises unicaracterística foram testados tamanhos de cadeia de 300.000, 500.000 e 800.000 iterações, períodos de descarte inicial (*burn-in*) de 1.000, 10.000, 30.000, 50.000 e 100.000 iterações. A melhor proposta para o tamanho de cadeia, *burn-in* e amostragem foi utilizada para a análise bicaracterística.

Um período de *burn-in* adequado é necessário para que sejam descartadas as estimativas iniciais que apresentam grande amplitude cujos valores interferem nas estimativas do valor médio dos componentes de variância.

Os períodos de *burn-in* foram testados pelo programa GIBANAL (VAN KAAM, 1998) cujas análises fornecem um intervalo de amostragem para cada *burn-in* testado. O critério de escolha do intervalo de amostragem foi feito de acordo com a correlação serial, resultante do *burn-in* e do processo de amostragem. Adicionalmente foram realizadas inspeções gráficas das cadeias para verificar se o *burn-in* atingiu a fase estacionária (fase do processo iterativo em que os valores estimados não apresentam grande amplitude) das mesmas adequadamente.

A correlação serial indica o grau de informação fornecido por cada amostra para a média *a posteriori*, refere-se à parte da medição das amostras relacionada ao tempo de variação do efeito estocástico causador de correlação entre observações dentro de um curto intervalo de tempo, tal efeito é menos correlacionado quando as medições forem mais espaçadas. Uma correlação alta indica redundância de informação, assim deseja-se que amostragem feita forneça a menor correlação possível.

A razão de se obter amostras ao invés da utilização da cadeia completa se deve ao fato facilitar a manipulação e o armazenamento das estimativas obtidas do grande volume soluções resultantes das cadeias. Assim a obtenção de uma amostra representativa adequada da cadeia pode contornar este problema.

A partir das amostras obtidas, foram calculadas as médias *a posteriori* para os componentes de (co)variância, parâmetros genéticos, regiões de alta densidade e o desvio-padrão associado a cada média. As regiões de alta densidade foram calculadas da seguinte maneira: as amostras obtidas pelo Amostrador de Gibbs foram ordenadas de forma crescente, em seguida foi calculado o intervalo entre as maiores e as menores amostras de modo que cada intervalo contivesse 95% das amostras. O menor intervalo foi escolhido para a apresentação e pode ser denominado intervalo de alta densidade representando o menor intervalo contendo 95% das amostras obtidas.

### 3. RESULTADOS E DISCUSSÃO

Nas Tabelas 2 a 7 são apresentados as correlações seriais ( $r$ ), intervalos de amostragem (skip), tamanhos de amostra, médias a *posteriori* da cadeia antes do *burn-in*, após o *burn-in* e após a amostragem da variância genética aditiva das análises unicaracterística.

Tabela 2 – Correlações seriais ( $r$ ), intervalos de amostragem (skip), tamanhos de amostra, médias a *posteriori* da cadeia antes do *burn-in*, após o *burn-in* e após a amostragem da variância genética aditiva da cadeia de 300.000 iterações para produção de leite.

<i>burn-in</i>	sem <i>burn-in</i>	após <i>burn-in</i>		após <i>burn-in</i> e amostragem			
	média	média	$r$	Skip**	$r$	média	amostra
<b>102*</b>	656.364,6	656.398,2	0,993	528	0,0606	658.468,6	568
1.000	656.364,6	656.459,8	0,993	566	-0,0196	659.580,9	566
10.000	656.364,6	656.471,9	0,993	487	0,0702	654.342,6	566
30.000	656.364,6	656.823,4	0,993	474	-0,0114	655.132	570
50.000	656.364,6	656.386,6	0,993	462	0,0939	656.275,9	542
100.000	656.364,6	654.736,3	0,993	432	0,0499	649.268,8	463

\* *burn-in* sugerido pelo GIBANAL

\*\* sugerido pelo GIBANAL

Tabela 3 – Correlações seriais (r), intervalos de amostragem (skip), tamanhos de amostra, médias a *posteriori* da cadeia antes do *burn-in*, após o *burn-in* e após a amostragem da variância genética aditiva da cadeia de 500.000 iterações para produção de leite.

<i>burn-in</i>	sem <i>burn-in</i>	após <i>burn-in</i>		após <i>burn-in</i> e amostragem			
	média	média	r	Skip**	r	média	amostra
<b>102*</b>	65.7102,4	657.121,9	0,993	556	0,057	659.879,4	900
1.000	65.7102,4	657.156,2	0,993	557	0,069	65.9047,3	896
10.000	65.7102,4	657.268,8	0,993	530	0,0038	657.540,8	925
30.000	65.7102,4	657.342,1	0,993	523	0,0402	658.238,7	899
50.000	65.7102,4	657.232,9	0,993	518	0,0575	659.484,6	869
100.000	65.7102,4	656.487,5	0,993	512	0,091	659.130,9	782

\* *burn-in* sugerido pelo GIBANAL

\*\* sugerido pelo GIBANAL

Tabela 4 – Correlações seriais (r), intervalos de amostragem (skip), tamanhos de amostra, médias a *posteriori* da cadeia antes do *burn-in*, após o *burn-in* e após a amostragem da variância genética aditiva da cadeia de 800.000 iterações para produção de leite.

<i>burn-in</i>	sem <i>burn-in</i>	após <i>burn-in</i>		após <i>burn-in</i> e amostragem			
	média	média	r	Skip**	r	média	amostra
<b>99*</b>	656.558,2	656.568,8	0,993	599	0,0443	656.480,6	1.336
1.000	656.558,2	656.580,9	0,993	600	0,0241	656.701,3	1.332
10.000	656.558,2	656.555,8	0,993	582	0,0194	657.611,6	1.358
30.000	656.558,2	656.631,4	0,993	577	0,0583	654.759,3	1.335
50.000	656.558,2	656.518,9	0,993	575	0,0607	656.311,4	1.305
100.000	656.558,2	656.312,3	0,993	577	0,0458	653.539,2	1.214

\* *burn-in* sugerido pelo GIBANAL

\*\* sugerido pelo GIBANAL

Tabela 5 – Correlações seriais (r), intervalos de amostragem (skip), tamanhos de amostra, médias a *posteriori* da cadeia antes do *burn-in*, após o *burn-in* e após a amostragem da variância genética aditiva da cadeia de 300.000 iterações para produção de gordura.

<i>burn-in</i>	sem <i>burn-in</i>	após <i>burn-in</i>		após <i>burn-in</i> e amostragem			
	média	média	r	Skip**	r	média	amostra
<b>99*</b>	968,421	968,482	0,993	553	0,0479	975,563	543
1.000	968,421	968,599	0,993	555	0,0602	974,928	539
1.0000	968,421	968,924	0,993	515	0,0328	965,072	564
30.000	968,421	969,357	0,993	500	0,0995	970,149	541
50.000	968,421	969,106	0,993	486	0,0976	968,031	515
100.000	968,421	967,035	0,993	462	0,0639	967,172	433

\* *burn-in* sugerido pelo GIBANAL

\*\* sugerido pelo GIBANAL

Tabela 6 – Correlações seriais (r), intervalos de amostragem (skip), tamanhos de amostra, médias a *posteriori* da cadeia antes do *burn-in*, após o *burn-in* e após a amostragem da variância genética aditiva da cadeia de 500.000 iterações para produção de gordura.

<i>burn-in</i>	sem <i>burn-in</i>	após <i>burn-in</i>		após <i>burn-in</i> e amostragem			
	média	média	r	Skip**	r	média	amostra
<b>99*</b>	970,591	970,625	0,993	582	0,0809	970,36	859
1.000	970,591	970,675	0,993	583	0,0609	976,396	856
10.000	970,591	970,919	0,993	556	0,0841	971,823	882
30.000	970,591	971,352	0,993	546	0,0317	975,621	861
50.000	970,591	971,276	0,993	541	0,0472	969,169	832
100.000	970,591	970,463	0,993	543	0,0512	964,316	737

\* *burn-in* sugerido pelo GIBANAL

\*\* sugerido pelo GIBANAL

Tabela 7 – Correlações seriais (r), intervalos de amostragem (skip), tamanhos de amostra, médias a *posteriori* da cadeia antes do *burn-in*, após o *burn-in* e após a amostragem da variância genética aditiva da cadeia de 800.000 iterações para produção de gordura.

<i>burn-in</i>	sem <i>burn-in</i>	após <i>burn-in</i>		após <i>burn-in</i> e amostragem			
	média	média	r	Skip**	r	média	amostra
<b>124*</b>	969,160	969,183	0,993	636	0,077	972,134	1.258
1.000	969,160	969,213	0,993	637	0,0541	970,850	1.255
10.000	969,160	969,354	0,993	614	0,0304	975,312	1.287
30.000	969,160	969,560	0,993	605	0,0379	969,930	1.273
50.000	969,160	969,755	0,993	602	0,0288	968,119	1.246
100.000	969,160	969,120	0,993	611	0,0826	972,799	1.146

\* *burn-in* sugerido pelo GIBANAL

\*\* sugerido pelo GIBANAL

Observa-se, de maneira geral, que todas as opções testadas para ambas as características apresentaram correlação serial abaixo de 0,1 (10%) após o *burn-in* e o intervalo de amostragem.

De acordo com os resultados conclui-se que, pela correlação serial, o melhor resultado para a produção de leite é a cadeia de 500.000 iterações com *burn-in* de 10.000 e intervalo de amostragem de 530.

Para produção de gordura, apesar do melhor resultado ter sido obtido com a cadeia de 800.000 iterações, *burn-in* de 50.000 e intervalo de amostragem de 602, optou-se por escolher a cadeia de 500.000 com *burn-in* de 30.000 e intervalo de amostragem de 546 para a análise bicaracterística uma vez que não houve aumento expressivo na correlação serial. Além disso, a escolha de análises com cadeias de 500.000 iterações ao invés de 800.000 são mais vantajosas com relação ao tempo gasto para serem processadas.

Na Tabela 8 são apresentadas as médias posteriores das herdabilidades ( $h^2$ ) com respectivos desvios-padrão e regiões de alta densidade (RAD) das melhores opções para ambas as características.

Tabela 8 – Estimativas das médias posteriores das herdabilidades ( $h^2$ ) com os respectivos desvios-padrão e regiões de alta densidade (RAD) das melhores opções para produção de leite (PL) e gordura (PG).

	$h^2$ (AD)	DP	RAD
PL	0,33	0,05	0,17
PG	0,33	0,05	0,17

As Figuras 1 e 2 apresentam os gráficos das estimativas da variância aditiva e residual após o *burn-in* nas melhores opções para as respectivas características.

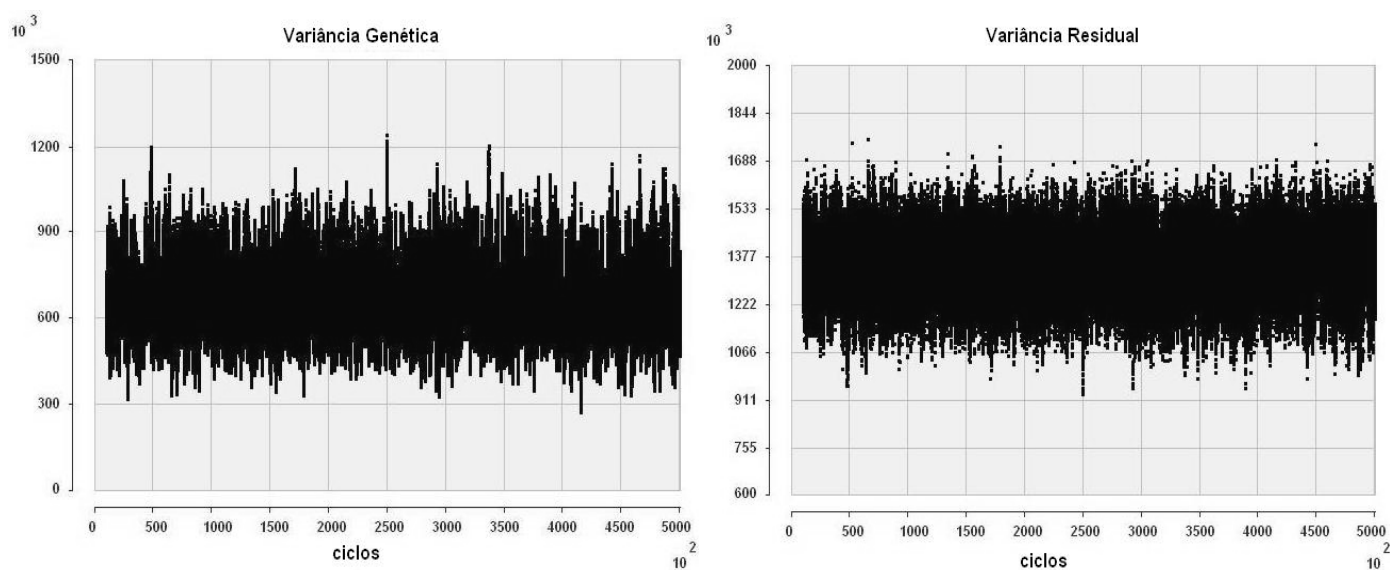


Figura 1 – Gráfico das amostras das estimativas da variância aditiva e residual em função do número de ciclos para a produção de leite após o descarte inicial de 10.000 iterações.

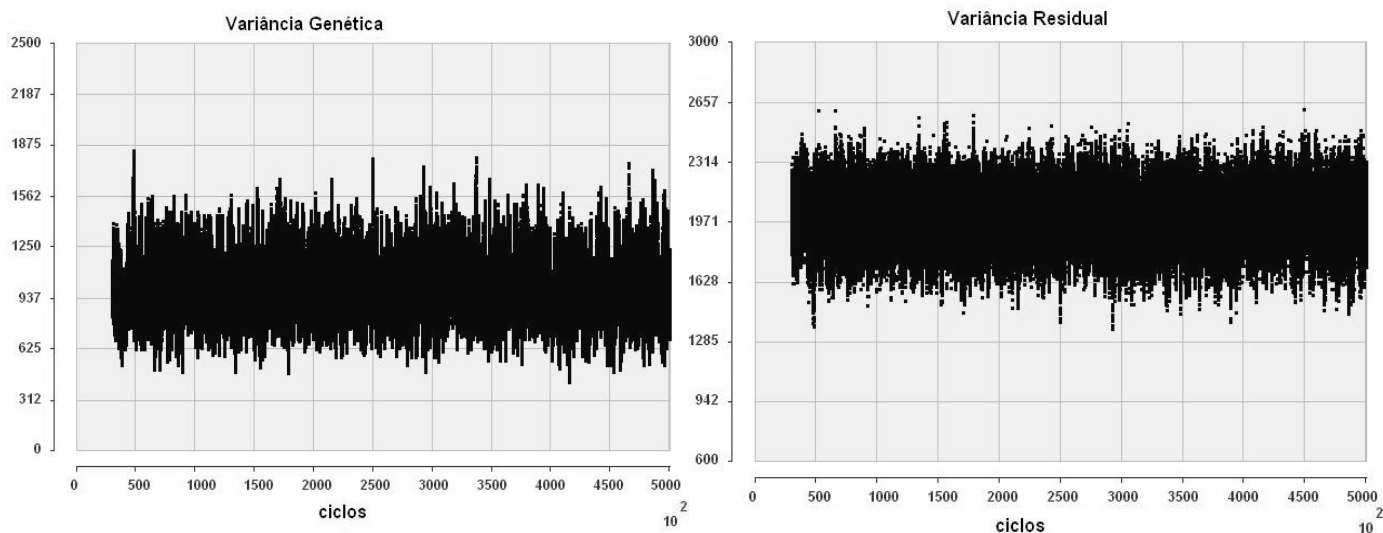


Figura 2 – Gráfico das amostras das estimativas da variância aditiva e residual em função do número de ciclos para a produção de gordura após o descarte inicial de 30.000 iterações.

Observa-se pelos gráficos que a convergência foi atingida em ambas as características.

Para a análise bicaracterística, foi necessário determinar um mesmo *burn-in* e intervalo de amostragem, adequados para as duas características sem favorecer o aumento da correlação serial para as mesmas. Para isso, foi escolhida a opção da análise para a produção de gordura que possuía o maior *burn-in* (30.000) menor intervalo de amostragem (546). Assim, foi testada esta opção para a produção de leite resultando em uma correlação serial de 0,036, valor inferior ao obtido pela sugestão de amostragem do GIBANAL. Adicionalmente não foram encontradas diferenças nos valores de herdabilidade, desvio-padrão e a região de alta densidade sugerindo que esta opção também pode ser recomendada para a análise da produção de leite.

Na Tabela 9 são apresentados os resultados obtidos na análise bicaracterística.

Tabela 9 – Estimativas das médias posteriores da (co)variância genética aditiva, residual, herdabilidade ( $h^2$ ), correlação genética ( $r_G$ ) e respectivos desvios-padrão (DP) e regiões de alta densidade (RAD) para a produção de leite (PL) e gordura (PG).

	$Va^1 \pm DP$	$Ve^2 \pm DP$	$h^2 \pm DP$ (RAD)
<b>PL</b>	660.872,1 ± 116.092,0	1.340.134,1 ± 90320,8	0,33 ± 0,05 (0,17)
<b>PG</b>	973,504 ± 176,777	1.987,408 ± 137,518	0,33 ± 0,05 (0,17)
<b>COV<sub>A</sub><sup>3</sup> ± DP</b>		24.019,50 ± 4.408,27	
<b>COV<sub>E</sub><sup>4</sup> ± DP</b>		48.407,07 ± 3.421,67	
<b>r<sub>G</sub> ± DP (RAD)</b>		0,95 ± 0,01 (0,17)	

1 – Variância genética aditiva; 2 – Variância residual; 3 – Covariância genética aditiva; 4 – Covariância residual.

Não houve diferenças nas estimativas entre as análises unicaracterística e a bicaracterística quanto à herdabilidade, desvios-padrão e a região de alta densidade.

As estimativas de herdabilidade obtidas para produção de leite estão dentro da amplitude encontrada por FALCÃO *et al.* (2006), em um estudo com vacas da raça holandesa utilizando métodos Bayesianos via Amostrador de Gibbs. Por outro lado, FIKSE *et al.* (2003) estimaram herdabilidade para produção de leite de  $0,32 \pm 0,032$  em animais da raça Guernsey em cinco diferentes países, utilizando uma cadeia com 150.000 iterações.

Resultados similares foram obtidos por REKAYA *et al.* (2003) e IBAÑEZ *et al.* (1999) em estudo de estimação Bayesiana de componentes de variância para produção de leite com dados de primeira lactação de animais da raça holandesa nos EUA e Espanha. Segundo PENA *et al.* (2001), altas estimativas de herdabilidade podem ser explicadas pela inclusão de apenas informações de primeira lactação.

As correlações seriais na análise bicaracterística foram maiores em comparação às obtidas nas análises unicaracterística, ainda assim, são consideradas de baixa magnitude variando de 0,095 a 0,132 como apresentadas na tabela 10.

Tabela 10 – Correlações seriais das estimativas dos componentes de (co)variância após a amostragem

	<b>Correlação Serial</b>
<b>Variância Aditiva (Produção de Leite)</b>	0,095
<b>Variância Residual (Produção de Leite)</b>	0,117
<b>Variância Aditiva (Produção de Gordura)</b>	0,108
<b>Variância Residual (Produção de Gordura)</b>	0,125
<b>Covariância Genética</b>	0,107
<b>Covariância Residual</b>	0,132

Na Figura 3 são apresentados os gráficos referentes às (co)variâncias genéticas aditivas e residuais em função do número de ciclos para a produção de leite e gordura após a retirada de amostras a cada 546 ciclos.

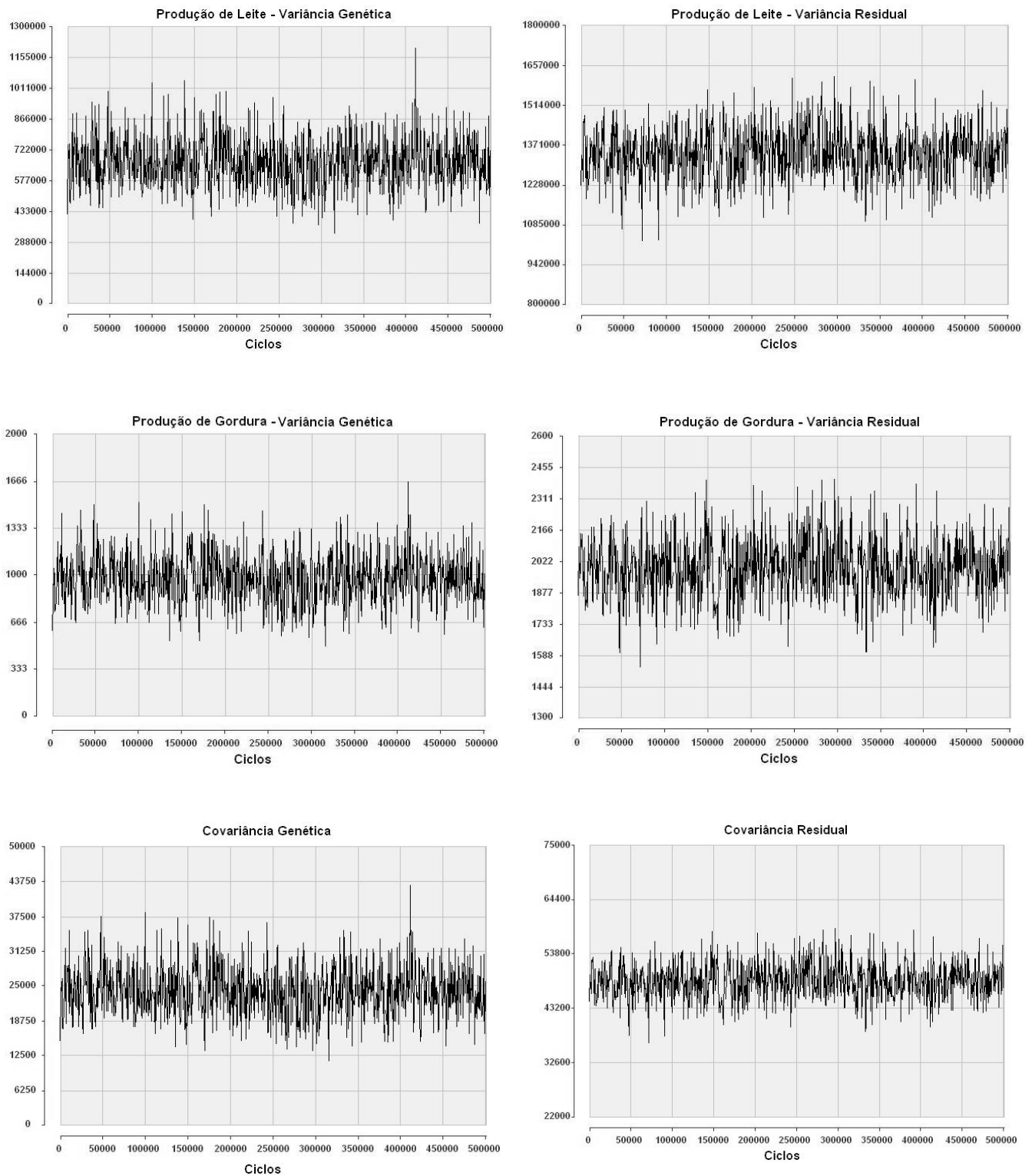


Figura 3 – Gráfico das estimativas das (co)variâncias genéticas aditivas e residuais em função do número de ciclos para a produção de leite e gordura após a retirada de amostras a cada 546 ciclos.

Pode-se observar em todos os gráficos que a cadeia atingiu a fase estacionária.

## **4. CONCLUSÃO**

A fase estacionária foi adequadamente atingida com uma cadeia de 500.000 iterações e descarte inicial de 30.000 iterações tanto nas análises uni quanto bicaracterística.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

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.

FALCAO, A. J. S., MARTINS, E. N., COSTA, C. N. *et al.* Heterocedasticidade entre estados para produção de leite em vacas da raça Holandesa, usando métodos Bayesianos via amostrador de Gibbs. *R. Bras. Zootec.*, mar./abr. 2006, vol.35, no.2, p.405-414. ISSN 1516-3598.

FIKSE W. F., REKAYA R. and WEIGEL, K. A. Genotype x Environment interaction for milk production in Guernsey cattle. *J. Dairy Sci.*, V.86, p.1821–1827, 2003

GERMAN, S.; GERMAN, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Transactions on Pattern Analysis and Machine Intelligence*. v.6, p.721-741, 1984.

GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theories. *J. Anim. Sci.*, v.63, p.217-244, 1986.

GIANOLA, D.; FOULLEY, J.L. Non linear prediction of latent genetic liability with binary expression: An empirical Bayes approach. In: *Proc. 2nd World Congr. Genet. Appl. Livest. Prod.*, Madrid, Spain. v.7, p.293-303, 1982.

HASTINGS, W.K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57:97-109.

IBAÑEZ, M. A., CARABAÑO, M. J., ALENDA R. Identification of sources of heterogeneous residual and genetic variances in milk yield data from Spanish

Holstein-Friesian population and impact on genetic evaluation. *Livestock Production Science*, n.59, p.33-49, 1999.

PENA, J. IBAÑEZ, M. A., CARABAÑO, M. J., JANSS, L. L. G. New Genetic Parameters for National Evaluations of Production Traits in Spanish Holsteins Excluding Selected base Animals from the Estimation of Genetic Variance. 2001. Interbull Technical Workshop in Verden, Germany, October 22-23 2000

REKAYA R., WEIGEL, K. A. and GIANOLA, D. Bayesian estimation of parameters of a structural model for genetic covariances between milk yield in five regions of the United States. *J. Anim. Sci.*, v. 86,p.1837–1844, 2003

VAN TASSELL, C.P.; VAN VLECK, L.D. A manual for use of MTGSAM. A set of FORTRAN programs to apply Gibbs sampling to animal models for variance components estimation (DRAFT). Lincoln: Department of Agriculture, Agricultural Research Service, 1995, 86p.

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.

WANG, C.S.; GIANOLA, D.; SORENSEN, D.A.; JENSEN, J.; CHRISTENSEN, A.; RUTLEDGE, J.J. Response to selection for litter size in Danish Landrace Pigs: A Bayesian analysis. *Theor. Appl. Genet.*, v.88, p.220-230, 1994.

## CONSIDERAÇÕES FINAIS

A idade da vaca ao parto influenciou linearmente a produção de leite e gordura, enquanto que a duração da lactação contribuiu negativamente para as estimativas de herdabilidade. Os menores valores das estimativas de herdabilidades comparados à literatura devem-se, possivelmente à ausência de ajustes dos dados de produção.

A alta correlação entre a produção de leite e gordura obtida sugere que parte dos genes que atuam na produção de leite também responde pela produção de gordura, de tal forma que a seleção para a produção de leite resulta, indiretamente, em aumentos na produção de gordura.

No estudo feito por abordagem Bayesiana, a fase estacionária foi adequadamente atingida com uma cadeia de 500.000 iterações e descarte inicial de 30.000 iterações tanto nas análises uni quanto bicaracterística.

As estimativas de herdabilidade foram similares às encontradas na literatura em estudos com informações de primeira lactação utilizando inferência Bayesiana, entretanto, foram maiores com relação às estimativas obtidas pela máxima verossimilhança restrita. A diferença nas estimativas das duas metodologias pode ser explicada pela utilização de um banco de dados relativamente pequeno. Uma opção para se testar esta hipótese seria a utilização de dados simulados, semelhantes aos deste estudo. Além disso, estudos adicionais com maior número de informações devem ser realizados com o objetivo de obter os componentes de variância com maior acurácia.

Ambas as escolas de inferência Bayesiana e freqüentista estão estabelecidas, e agora, nenhuma delas apresenta dificuldades operacionais, com exceção de alguns casos complexos. Existem softwares disponíveis para analisar uma grande variedade de problemas para os dois pontos de vista. A escolha de uma escola ou outra em determinada ocasião deve-se basear nas soluções que uma escola pode oferecer e a outra não, em quão fácil o

problema será resolvido e ainda na conveniência do pesquisador quanto à maneira de conduzir seus resultados.