

Universidade Federal de Alfenas

Andréia do Carmo de Oliveira

Modelos mistos aplicados à variação genética do perfil
de ácidos graxos e sólidos totais do leite

Alfenas - MG

2021

Andréia do Carmo de Oliveira

Modelos mistos aplicados à variação genética do perfil
de ácidos graxos e sólidos totais do leite

Trabalho de Conclusão de Curso apresentado à Universidade Federal de Alfenas como parte dos requisitos para obtenção da Licenciatura em Matemática.

Área de concentração: Estatística Aplicada.

Orientador: Prof. Dr. Denismar Alves Nogueira.

Coorientadora: Prof.^a Dr.^a Juliana Petrini.

Alfenas - MG

2021

Andréia do Carmo de Oliveira

Modelos mistos aplicados à variação genética do perfil de ácidos graxos e sólidos totais do leite

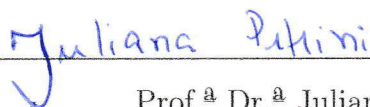
A banca examinadora abaixo-assinada, aprova o Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do Certificado de Conclusão do Curso de Licenciatura em Matemática pela Universidade Federal de Alfenas.

Aprovado em: 02 de março de 2021.



Prof. Dr. Denismar Alves Nogueira

Orientador



Prof.^a Dr.^a Juliana Petrini

Coorientadora



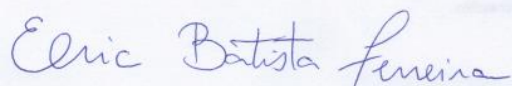
Prof.^a Dr.^a Natália da Silva Martins Fonseca

Universidade Federal de Alfenas



Prof.^a Dr.^a Mary Ana Petersen Rodriguez

Universidade Estadual de Minas Gerais



Prof. Dr. Eric Batista Ferreira (Suplente)

Dedico este trabalho aos meus pais José Reinaldo e Ivonete, aos meus irmãos Letícia e Felipe, aos meus sobrinhos Miguel e Lorena, ao meu esposo Reginaldo e ao meu amor incondicional, meu filho Augusto, que sempre estiveram ao meu lado, me apoiaram, acreditaram no meu sonho e ajudaram a torná-lo realidade.

Agradecimentos

A Deus pelo dom da vida e por sempre iluminar meu caminho, dando-me forças para vencer todos os obstáculos e tornar esse sonho possível.

Aos meus pais José Reinaldo e Ivonete, pelo amor incondicional, pela educação, pela confiança e que apesar de todas as dificuldades ao longo do caminho, não mediram esforços para me apoiar e me ver feliz.

Aos meus irmãos Letícia e Felipe, pela amizade, confiança e por sempre estarem dispostos a me ajudar.

Ao meu filho Augusto e meus sobrinhos Miguel e Lorena por serem minha força e inspiração. Amo vocês incondicionalmente.

Ao meu amigo Thomás pelo companheirismo e amizade, tanto nas horas de estudo quanto nas horas de lazer. Serei eternamente grata pelos momentos que passamos juntos.

A minha amiga Ludmila por todos os momentos que passamos juntas desde o início da caminhada acadêmica. Agradeço pelo carinho, companheirismo, amizade e por sempre estar ao meu lado nos momentos tristes e alegres.

A minha irmã de coração Michele Martins, por todos os momentos que passamos juntas desde a nossa viagem para Gramado-RS. Sou muito grata pelos inúmeros conselhos e por sempre me incentivar nas horas mais difíceis da minha vida. Jamais esquecerei tudo que fez por mim.

Aos amigos que fiz no decorrer dos semestres, em especial a Amanda, Thales e Mariane. Os momentos que passamos juntos desde o desenvolvimento do projeto de extensão ficarão marcados. Agradeço pela amizade, companheirismo, convivência e momentos de puro agito durante esses anos. Vocês são presentes que a Matemática me deu.

Aos meus orientadores Juliana Petrini e Denismar Nogueira, pela paciência, ensinamentos, orientação, força e amizade durante o desenvolvimento do trabalho.

Ao Prof. Dr. Fabrício Avelar, por ter me guiado nos primeiros passos dessa longa caminhada, pelo exemplo de pessoa, pelo profissionalismo, pela inspiração, por me mostrar a beleza da Matemática, por todo carinho, incentivo e conhecimento compartilhado. Com certeza, seu apoio e orientação foram fundamentais para escolha da minha formação profissional.

A Prof. Luciana Borges, pela amizade, atenção, carinho e pelos sábios ensinamentos, responsáveis pelo meu crescimento pessoal e profissional.

Aos meus colegas da Matemática, em especial aos amigos Hernando, Natã, Paulinha, Reginaldo, Diego, Valdeline, Vinícius, Roberta, Ronaldo e Edson por tornarem a vida acadêmica muito mais divertida.

Aos professores do Departamento de Matemática da UNIFAL-MG que fizeram parte da minha formação, em especial ao José Carlos, Andréa Cardoso, José Paulo, Anderson e Rejane.

Um agradecimento especial eu reservo, ao meu esposo Reginaldo Leandro da Silva, o seu amor, força e carinho, sempre me fizeram reagir e lembrar que vale a pena continuar na luta mesmo que os obstáculos pareçam invencíveis.

A todos que de alguma forma foram responsáveis por essa vitória tão esperada ao longo dos anos, meus sinceros agradecimentos.

“O êxito da vida não se mede pelo caminho que você conquistou, mas sim pelas dificuldades que superou no caminho.”

(Abraham Lincoln)

Resumo

Modelos lineares mistos podem ser definidos como modelos que incluem efeitos fixos e aleatórios adicionalmente ao efeito do resíduo e à média. No melhoramento genético animal, os efeitos genéticos são considerados como efeitos aleatórios e o modelo misto é utilizado para a predição do mérito genético dos animais e estimação da variância do componente genético. Com a recente disponibilidade das informações genômicas do animal, tornou-se ainda possível expandir e aumentar a acurácia de predição por meio do aumento do número e qualidade das informações consideradas. Dessa forma, o objetivo neste estudo foi utilizar o modelo misto visando à estimação da variância aditiva e predição dos efeitos aleatórios para características de produção de bovinos leiteiros. Para tanto, foram utilizados 62.740 dados de produção de leite em (kg/dia), e porcentagem de sólidos totais, ácidos graxos saturados e insaturados no leite de 6.022 vacas da raça Holandesa. Adicionalmente, foram utilizadas as informações do genótipo de 1.152 dessas vacas. Os componentes de variância bem como as predições dos efeitos aleatórios foram obtidos por meio de um modelo misto sob o método da máxima verossimilhança restrita com o algoritmo AI (*average information*), considerando o efeito genético como aleatório. Coeficientes de herdabilidade moderados foram estimados para teor de sólidos totais e ácidos graxos saturados, sendo 0,28 e 0,24 respectivamente. Acurácias altas (maiores que 0,70) foram obtidas para os valores preditos para teor de sólidos totais e ácidos graxos saturados, sendo 0,758 e 0,703, respectivamente. Ganhos genéticos e tendências genéticas favoráveis foram verificados para produção de leite e teor de sólidos totais. Verificou-se que existe variação genética para a característica de sólidos totais, indicando o uso dessa característica como critério de seleção em programas de melhoramento animal, destacando-se o fato de que modelos lineares mistos são adequados na estimação dos componentes de variância das características estudadas.

Palavras-chave: Ácidos Graxos. Componentes de Variância. Genômica. Efeitos Aleatórios.

Abstract

Linear mixed models can be defined as models that include both fixed and random effects in addition to the intercept and residual. In animal breeding, genetic effects are considered as random effects and mixed models are used to predict the genetic merit of the animals under selection and to estimate the genetic variance. The recent availability of genomic information allowed the increase of prediction accuracy by increasing the amount of information and by improving its quality. This way, the aim of this study was to use a mixed model to estimate the additive genetic variance and to predict the random effects for milk production traits. A total of 62,740 measures of milk yield (kg/day), and percentage of total solids, saturated and unsaturated fatty acids in milk from 6,022 Holstein cows was used. In addition, it was used the genotype information from 1,152 of these cows. Variance component estimates as well as the random effect predictions were obtained through a mixed model under restricted maximum likelihood method with average information algorithm, assuming the genetic effect as random. Moderate heritability coefficients were estimated for total solids content and saturated fatty acids, with values of 0.28 and 0.24, respectively. High accuracies (higher than 0.70) were obtained for breeding values for total solids and saturated fatty acids, with mean values of 0.758 and 0.703, respectively. Favorable genetic gain and genetic trends were verified for milk production and total solids. There was genetic variation for total solids, which allows the use of this trait as a selection criterion in animal breeding programs. Moreover, we highlight the adequate use of mixed models for the genetic evaluation of these studied traits.

Keywords: Fatty acids. Genomics. Random effects. Variance components.

Sumário

1 INTRODUÇÃO	11
2 REVISÃO DE LITERATURA	13
2.1 Melhoria Genética na Bovinocultura Leiteira	13
2.1.1 Objetivos e Critérios de Seleção	13
2.1.2 Componentes de Variância	15
2.1.3 Coeficiente de Herdabilidade	16
2.1.4 Avaliação Genética	16
2.2 Modelos Mistos	17
2.2.1 Estimação de β e Predição de u	19
2.2.2 Estimação dos Parâmetros das Matrizes de Covariância R e G	21
2.2.3 Métodos Iterativos	22
2.3 Modelo Misto no Melhoramento Genético Animal	24
2.4 Metodologias Estatísticas em Seleção Genômica	24
2.4.1 BLUP Genômico (GBLUP)	24
2.4.2 <i>single-step</i> GBLUP	25
2.4.3 Construção da Matrix H	26
3 MATERIAIS E MÉTODOS	27
3.1 Animais e Dados	27
3.2 Genótipos	28
3.3 Análise Genética	28
3.3.1 Estimação dos Componentes de Variância	28
3.3.2 Predição dos Valores Genéticos	29
4 RESULTADOS E DISCUSSÃO	30
4.1 Estatísticas Descritivas	30
4.2 Parâmetros Genéticos	31
4.3 Valores Genéticos Preditos e Acurácia	34
4.4 Ganho Genético e Tendência Genética	35
5 CONCLUSÕES	38

1 INTRODUÇÃO

O objetivo do melhoramento genético animal é alterar geneticamente uma população, isto é, aumentar a frequência dos genótipos favoráveis às características de interesse econômico. Isso é predominantemente realizado por meio do uso de animais geneticamente superiores na reprodução. No entanto, esse processo, denominado de seleção genética, depende da estimação da variação observada na característica que é oriunda da ação dos fatores genéticos (variância genética), da predição do mérito genético dos indivíduos sob seleção (valores genéticos) bem como da quantificação da proporção da variação total (variância fenotípica) que corresponde ao componente genético (herdabilidade) (FERRAZ, ELLER, 2010).

Nesse contexto, o modelo linear misto tem-se apresentado como metodologia adequada à avaliação genética em programas de melhoramento genético. Neste modelo, o efeito genético é assumido como um efeito aleatório, de forma que os valores genéticos são obtidos com propriedades de melhor preditor linear não viesado (*best linear unbiased predictor*, BLUP) (HENDERSON, 1975). Além disso, este modelo apresenta outras propriedades ótimas, tais como a maximização da correlação entre os valores genéticos verdadeiros e preditos; a estimação e predição não viesados em um método único; a predição do mérito genético de indivíduos observados e não observados; além da minimização das variâncias e erros associados (GARCIA, 2004; WHITE; HODGE, 1989). As predições resultantes do modelo misto permitem ainda comparações entre indivíduos na população que diferem quanto ao ambiente ao qual estão submetidos e quanto à quantidade de informação fenotípica disponível sobre eles e seus parentes. Assim, por exemplo, as tendências genéticas ao longo dos anos e gerações podem ser estimadas livres de mudanças ambientais (HENDERSON et al. 1959).

Atualmente, com a crescente demanda por sistemas de produção mais eficientes associados a produtos de qualidade nutricional, os objetivos dos programas de melhoramento genético têm-se ampliado, podendo incluir caracteres relacionados à composição do leite, como o teor de sólidos totais (gordura, lactose e proteína) e perfil de ácidos graxos no leite. No entanto, algumas dessas características (principalmente os ácidos graxos) não são intensivamente mensuradas em rebanhos leiteiros, visto que essa mensuração pode ser cara e exigir reagentes específicos e mão-de-obra especializada. Com isso, a possibilidade de se alterar a composição do leite por meio do melhoramento genético é afetada, dado

que o número limitado de fenótipos disponíveis pode restringir a acurácia da predição dos valores genéticos (PETRINI et al., 2019).

Contudo, na última década, tornou-se disponível à avaliação genética as informações de centenas a milhares de marcadores genéticos, os quais são variações nas sequências do DNA que podem estar associadas às características de interesse. Dessa forma, tem sido possível aumentar a acurácia da predição em modelos mistos, inclusive permitindo a predição acurada para fenótipos de cara e/ou difícil mensuração, e com número reduzido de fenótipos (KRAG et al., 2013; van EENENNAAM et al., 2014). Na denominada predição genômica (ou seleção genômica), as informações dos marcadores genéticos são utilizadas na estimação do grau de relacionamento entre os indivíduos, o qual é considerado na construção da matriz de variâncias e covariâncias associada ao efeito aleatório de animal (de los CAMPOS et al., 2013).

O uso de informações genômicas na avaliação genética tem ocasionado o aumento do ganho genético por meio do aumento da acurácia de predição associada à redução do intervalo de gerações, visto que permite a predição acurada de animais jovens, com pouca ou nenhuma informação fenotípica disponível acerca de si e suas progênes (van EENENNAAM et al., 2014).

Assim, o objetivo neste estudo é utilizar um modelo linear misto no estudo da variação genética da composição do leite (especificamente teor de ácidos graxos e sólidos totais), com a inclusão da informação de marcadores genéticos na análise.

A presente monografia encontra-se estruturada da seguinte forma:

Na primeira seção são apresentados conceitos e métodos que foram utilizados para o desenvolvimento deste trabalho. Nesta seção é visto a revisão de literatura sobre modelo linear misto, melhoramento genético e os métodos estatísticos em seleção genômica utilizados neste trabalho.

Na segunda seção são descritos a metodologia utilizada nas análises estatísticas e os dados utilizados no desenvolvimento deste trabalho.

Na terceira seção são apresentados os resultados e discussão do presente trabalho. Nesta seção constam as estatísticas descritivas, as estimativas dos componentes de variância, parâmetros genéticos, predição dos valores genéticos, ganho genético e tendência genética.

2 REVISÃO DE LITERATURA

2.1 Melhoramento Genético na Bovinocultura Leiteira

O setor leiteiro tem um importante papel na ordem econômica e social do agronegócio brasileiro, com uma participação significativa no PIB da pecuária (VILELA, RESENDE, 2014). Em 2019, a produção de leite no Brasil foi de 34,8 bilhões de litros, atingindo a marca de 2.142 litros de leite por vaca por ano (IBGE, 2019). Nos últimos anos, a atividade leiteira no Brasil, registrou aumento na produção de leite, mas com redução no número de produtores e menos vacas ordenhadas (ZOCCAL, 2020). Essa situação se deve ao aumento do profissionalismo na gestão da atividade e adoção de tecnologias por meio de assistência técnica especializada. Contudo, se comparado a países de tradição no setor pecuarista como Estados Unidos e Índia, o Brasil apresenta baixa produtividade, sendo necessário implantar medidas que visem aumentar a eficiência da produtividade leiteira no país, como o melhoramento genético.

O melhoramento genético pode ser definido como a ciência que estuda as ações da genética dos indivíduos e do ambiente na determinação de suas características de interesse econômico (FERRAZ; ELER, 2010). Ele é realizado mediante a exploração das diferenças genéticas existentes entre raças e linhagens e também entre indivíduos de uma mesma raça ou linhagem, com uso da seleção e dos sistemas de cruzamentos como principais ferramentas (MOURÃO; FORMIGONI, 2006). Especificamente a seleção consiste em escolher os animais que, pela união de seus gametas, formarão a próxima geração (FERRAZ; ELER, 2010). Dessa forma, o objetivo da seleção é permitir que os indivíduos com alelos favoráveis se reproduzam, de maneira que a próxima geração seja, em média, superior à anterior. Este procedimento tem como principal efeito a alteração das frequências alélicas nos *loci* que controlam a característica sob seleção (BOURDON, 2000).

2.1.1 Objetivos e Critérios de Seleção

Denomina-se objetivo de seleção como a combinação de características economicamente importantes de acordo com um determinado sistema de produção (CARDOSO et al., 2004). Por sua vez, o critério de seleção representa o conjunto de informações a respeito de uma característica ou conjunto de características em que a seleção se baseia, visando avaliar e ordenar os candidatos à seleção para a característica alvo do melhora-

mento (VAYEGO, 2007).

O leite bem como os derivados lácteos podem ser considerados alimentos balanceados e nutritivos, representando uma fração importante da dieta humana. O leite bovino é composto de 87,5% de água e 12,5% de sólidos totais, os quais correspondem ao conteúdo de 3,6% de gordura, 3,0% de caseína (proteína), 0,6% de albumina, 4,6% de lactose e 0,7% de minerais (GONZÁLEZ; NORO, 2011). Os sólidos totais são fonte de ácidos graxos, aminoácidos essenciais e cálcio (PEREIRA, 2014), e possuem valor industrial, visto a sua relação com a produção de derivados lácteos como queijos, manteiga e iogurtes (HARDING, 1995).

A quantidade e composição dos sólidos totais pode ser influenciada pelo período da lactação, idade da vaca, dieta e demais condições ambientais. Adicionalmente, existe variação quanto ao teor de sólidos totais entre e dentro de raças, de forma que o melhoramento genético apresenta-se como ferramenta importante em sistemas de produção de bovinos leiteiros (HARDING, 1995).

Nos últimos anos os consumidores têm demonstrado um grande interesse na qualidade do leite. Segundo Haug et al. (2007), a composição de gordura, ácido oléico, ácido linoléico conjugado, ácidos graxos e ômega-3 podem ser benéficos para a saúde humana. Este fato pode levar os consumidores a prestar uma atenção especial à qualidade da gordura, principalmente em termos de perfil de ácidos graxos. No decorrer dos séculos os hábitos alimentares do ser humano passaram por uma drástica mudança. Houve um significativo aumento na ingestão de gordura e nas taxas de ácidos graxos, favorecendo ômega-6, presente em óleos vegetais e reduzindo a ingestão de ômega-3, o qual é conhecido por suas propriedades antitrombóticas, antiinflamatórias e antiaterogênicas (SIMOPOULOS, 2016).

Segundo Bobbo et al. (2020), os lipídios do leite são compostos principalmente de triglicerídeos, os quais possuem uma molécula de glicerol e três moléculas de ácidos graxos. O alto número de ácidos graxos conjugados com glicerol, torna o leite um alimento complexo do ponto de vista do perfil lipídico (PARODI, 2004).

O componente gorduroso do leite pode ser influenciado por vários fatores, incluindo a raça do animal, a dieta, o sistema de produção e o estágio de lactação (Hanus et al. 2018). De acordo com Bobbo et al. (2020), mudanças na dieta dos animais pode ocasionar variações a curto prazo no perfil de ácidos graxos do leite, mas é possível aplicar

uma abordagem genética para melhorar a qualidade da gordura contida no leite. Essa estratégia permite melhorias a longo prazo na qualidade da gordura e tem por objetivo selecionar os animais geneticamente superiores afim de obter leite com perfil de ácidos graxos favorável a saúde humana, como por exemplo, rico em ômega-3.

O desempenho na produção de leite também é de grande relevância na atividade leiteira. Com o constante aumento da população mundial é preciso aumentar a produção de leite e seus derivados. Para isso, é necessário o aumento da produtividade e eficiência por animal, para que assim os produtores possam expandir a sua produção sem a necessidade de aumentar o rebanho.

A produção de leite, é considerada a principal característica de interesse econômico em raças leiteiras. Simm (1998) relatou em seus estudos que a seleção genética para produção de leite é sinônimo de maior lucratividade. Sendo assim, a produção de leite é um importante critério de seleção a ser inserido em programas de melhoramento genético de bovinos leiteiros.

2.1.2 Componentes de Variância

Os fenótipos de características produtivas são resultantes da ação de diversos genes, de fatores ambientais bem como da interação entre eles, de forma que a variação fenotípica pode ser dividida em variação genética, variação ambiental e variação devido à interação entre genótipo e ambiente (MACKAY, 2001). Especificamente a variação genética pode ser particionada em componentes aditivo, de dominância e epistático. A estimação dos componentes de variância aditivos e não aditivos utiliza a similaridade fenotípica entre indivíduos aparentados e a contribuição esperada dos efeitos aditivos e não aditivos a esta similaridade (FALCONER; MACKAY, 1996; LYNCH; WALSH, 1998).

A variância aditiva ($\hat{\sigma}_a^2$) em um único *locus* pode ser definida como a variância explicada pela regressão do valor genotípico (valor esperado da média fenotípica em cada classe genotípica) sobre o genótipo. Por sua vez, a variância de dominância ($\hat{\sigma}_d^2$) pode ser entendida como a variância genética residual não explicada por esta regressão. Seja $a = \frac{(\mu_{BB} - \mu_{AA})}{2}$ e $d = \mu_{AB} - \frac{(\mu_{AA} + \mu_{BB})}{2}$ com μ_{AA} , μ_{AB} e μ_{BB} sendo as médias fenotípicas nas classes genotípicas AA, AB e BB, respectivamente. Sob a pressuposição de equilíbrio de Hardy-Weinberg (HWE), $\hat{\sigma}_a^2$ é igual a $2p(1-p)[a + (1-2p)d]^2$, e $\hat{\sigma}_d^2$ é igual a $[2p(1-p)d]^2$ e a variância genotípica ($\hat{\sigma}_g^2$) é igual a $(\hat{\sigma}_a^2) + (\hat{\sigma}_d^2)$, com p sendo a frequência alélica de

B (FALCONER; MACKAY, 1996).

A $\widehat{\sigma}_a^2$ é, portanto, a variância do efeito médio de substituição de alélica (α), com $\alpha = +(1 - 2p)d$ (LYNCH; WALSH, 1998), o qual inclui o termo de dominância devido à interação entre os dois alelos. Nisto, a $\widehat{\sigma}_d^2$ é a variação nos desvios dos valores genotípicos da regressão (ZHU et al., 2015).

A partir das estimativas dos componentes de variância é possível estimar a herdabilidade. Esta é uma medida da contribuição da genética sobre o fenótipo (ZAITLEN; KRAFT, 2012).

2.1.3 Coeficiente de Herdabilidade

A herdabilidade (em seu sentido restrito) é usualmente definida como a proporção da variação fenotípica devida a efeitos genéticos aditivos. Trata-se de um parâmetro genético populacional que ajuda na compreensão da arquitetura genética de características complexas (LEE et al., 2011), determinando o grau de semelhança existente entre indivíduos aparentados. Adicionalmente, a herdabilidade possui um papel preditivo, expressando a confiança no valor fenotípico como preditor do valor genético (FALCONER; MACKAY, 1996).

A herdabilidade permite indicar o potencial para seleção fenotípica e melhorar geneticamente uma característica. Uma herdabilidade alta sugere que a seleção individual produzirá um rápido melhoramento genético, mas um baixo valor de herdabilidade indica que o progresso na seleção será lento e outros métodos deveriam ser adotados para melhorar a característica (FERREIRA, 2011).

A herdabilidade pode variar de 0 a 1 (0% a 100%). Segundo Pereira (2008), valores de herdabilidade de 0 a 0,15 são considerados baixos, entre 0,15 a 0,3, moderados e acima de 0,3 de alta herdabilidade.

2.1.4 Avaliação Genética

A avaliação genética de um animal consiste na predição do seu valor genético aditivo, isto é, do seu mérito genético como reprodutor. O valor genético de um indivíduo depende da ação dos alelos dos genes envolvidos na determinação do fenótipo de interesse, do número de informações referentes aos animais avaliados, do relacionamento entre estes animais, além dos efeitos ambientais e da acurácia com que estes são estimados. A predição

dos valores genéticos está ainda condicionada à estimação dos componentes de variância e principalmente à herdabilidade, visto que esta expressa a confiança do valor fenotípico como preditor do valor genético (FERRAZ; ELER, 2010).

Atualmente com a inclusão da informação genômica na avaliação genética, tem sido possível aumentar a acurácia dos valores genéticos preditos, permitindo a seleção de animais jovens com confiança alta. Dessa forma, reduz-se o intervalo de gerações e aumenta-se o ganho genético. Essas vantagens são particularmente aplicáveis na bovinocultura leiteira devido ao seu longo intervalo de gerações; fenótipos restritos a fêmeas; e características importantes, mas difíceis ou caras de serem mensuradas, como fertilidade, resistência a doenças e composição de ácidos graxos no leite. A adoção da seleção genômica em programas de melhoramento genético pode reduzir o intervalo de gerações em dois anos, aumentando a taxa de ganho genético de 60 a 120% (HAYES et al., 2013).

Comparativamente à avaliação genética tradicional, Wiggans et al. (2011) estimaram ganhos em acurácia entre 17,0 e 47,6% para características de produção de leite, produção de gordura e proteína, vida produtiva, escore de células somáticas e taxas de prenhez com a inclusão de informações genotípicas na avaliação de touros jovens da raça Holandesa. Mais recentemente, foi estimado nos Estados Unidos uma redução entre 5 a 65% no intervalo de gerações e um aumento de aproximadamente 15 a 46% no ganho genético após a introdução da informação genômica na avaliação genética (GARCÍA-RUIZ et al., 2016).

2.2 Modelos Mistos

Segundo Pinheiro e Bates (2000), modelo misto é aquele no qual estão presentes efeitos fixos e aleatórios, além do erro experimental e da constante μ (LITTELL, et al., 2006). Camarinha Filho (2002), denota o modelo linear misto da seguinte forma:

$$y_{ijk} = \mu + \beta_i + a_j + \varepsilon_{ijk} \quad (1)$$

em que, y_{ijk} é a observação referente à k -ésima repetição do nível i de uma fonte de efeitos fixos ao nível j de uma fonte de efeitos aleatórios; μ é uma constante inerente a todas as observações; β é o efeito do nível i do fator fixo, $i = 1, \dots, p$; a é o efeito do nível j , do fator aleatório, no nível i do fator fixo, $j = 1, \dots, q$; ε_{ijk} é o erro aleatório associado a observação

y_{ijk} .

O modelo misto pode ser especificado na forma matricial (HENDERSON, 1975):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (2)$$

em que \mathbf{y} representa o vetor das respostas individuais ou vetor de observações, de dimensão $n \times 1$; \mathbf{X} é uma matriz do planejamento experimental relativa aos efeitos fixos, de dimensão $n \times p$; $\boldsymbol{\beta}$ é o vetor de efeitos fixos, desconhecidos, de dimensão $p \times 1$; \mathbf{Z} é uma matriz conhecida do planejamento experimental, relativa aos efeitos aleatórios, de dimensões $n \times q$; \mathbf{u} é o vetor de efeitos aleatórios, desconhecidos, de dimensão $q \times 1$; $\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios não observáveis, de dimensão $n \times 1$.

No modelo matricial, assume-se que $E(\mathbf{u}) = 0$ e $E(\boldsymbol{\varepsilon}) = 0$ com matrizes de covariâncias representadas por:

$$\text{Var} = \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

com suposição que \mathbf{u} e $\boldsymbol{\varepsilon}$ são normalmente distribuídos e independentes, assim $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$ e $\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{R})$.

O modelo linear misto é comumente especificado em termos das respostas condicionais aos efeitos aleatórios, sendo dado por:

$$\mathbf{y}|\mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R}) \quad (3)$$

em que \mathbf{y} é um vetor característico de medidas para unidades experimentais.

Entretanto, as inferências são realizadas no modelo marginal, assim,

$$\begin{aligned} \text{Var}[\mathbf{y}] &= \text{Var}[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}] = \text{Var}[\mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}] \\ \text{Var}[\mathbf{y}] &= \text{Var}[\mathbf{Z}\mathbf{u}] + \text{Var}[\boldsymbol{\varepsilon}] + \text{Cov}[\mathbf{Z}\mathbf{u}, \boldsymbol{\varepsilon}] + \text{Cov}[\boldsymbol{\varepsilon}, \mathbf{Z}\mathbf{u}] \\ \text{pois, } \text{Var}[\mathbf{X}\boldsymbol{\beta}] &= \text{Cov}[\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\varepsilon}] = \text{Cov}[\boldsymbol{\varepsilon}, \mathbf{X}\boldsymbol{\beta}] = \Phi \\ \text{Var}[\mathbf{y}] &= \mathbf{ZV}[\mathbf{u}]\mathbf{Z}' + \mathbf{R} + \mathbf{ZCov}[\mathbf{u}, \boldsymbol{\varepsilon}] + \text{Cov}[\boldsymbol{\varepsilon}, \mathbf{u}]\mathbf{Z}' \end{aligned}$$

Com relação as propriedades do modelo matricial, tem-se que:

$$V = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} + \mathbf{Z}\Phi_{q \times n} + \Phi_{n \times q}\mathbf{Z}'$$

logo,

$$V = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

e

$$E[\mathbf{y}] = E[\mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}] = E[\mathbf{X}\beta] + E[\mathbf{Z}\mathbf{u}] + E[\boldsymbol{\varepsilon}] = \mathbf{X}\beta$$

Assim, $\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{V})$. Henderson (1975) propôs estimar β e prever \mathbf{u} por meio das equações de modelo misto, obtidas por meio da maximização da distribuição conjunta de \mathbf{u} e \mathbf{y} em relação à β e \mathbf{u} .

2.2.1 Estimação de β e Predição de u

Nesta classe de modelos, há o interesse prático de estimar os efeitos fixos, os componentes de variância e prever os efeitos aleatórios. A obtenção das equações de modelos mistos pode ser feita através da minimização da soma de quadrados dos resíduos ou pela maximização da função densidade de probabilidade conjunta de \mathbf{y} e \mathbf{u} (FREITAS, 2007).

A função densidade de probabilidade de \mathbf{y} é dada por:

$$f(\mathbf{y}) = \frac{1}{2\phi^{n/2}(|\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}|)^{1/2}} e^{-1/2[(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{X}\beta)]} \quad (4)$$

A função densidade de probabilidade conjunta de \mathbf{y} e \mathbf{u} pode ser escrita como:

$$f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u}) \cdot f(\mathbf{u}) \quad (5)$$

Assim,

$$f(\mathbf{y}, \mathbf{u}) = \frac{1}{(2\phi)^{n/2}(|\mathbf{R}|)^{1/2}} e^{-1/2[(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u})'(\mathbf{R})^{-1}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u})]} \cdot \frac{1}{(2\phi)^{j/2}(|\mathbf{G}|)^{1/2}} e^{-1/2[(\mathbf{u} - \mathbf{0})'(\mathbf{G})^{-1}(\mathbf{u} - \mathbf{0})]} \quad (6)$$

em que n e j são os postos das matrizes \mathbf{R} e \mathbf{G} , respectivamente.

Para maximizar a função densidade de probabilidade conjunta, pode-se usar a trans-

formação por logaritmo. O logaritmo neperiano da função de verossimilhança é:

$$l(\beta, \mathbf{u}|\mathbf{y}) = \frac{(n+j)}{2} \ln(2\phi) - \frac{1}{2}(\ln \mathbf{R} + \ln \mathbf{G}) - \frac{1}{2}(\mathbf{y}'\mathbf{R}^{-1}\mathbf{y} - 2\mathbf{y}'\mathbf{R}^{-1}\mathbf{X}\beta - 2\mathbf{y}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + 2\beta'\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\beta + \mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{u}'\mathbf{G}^{-1}\mathbf{u})$$

Derivando-se $l(\beta, \mathbf{u}|\mathbf{y})$ em relação a β e \mathbf{u} , e tomando-se as derivadas parciais iguais a $\mathbf{0}$, obtém-se:

$$\begin{bmatrix} \frac{\partial l(\beta, \mathbf{u}|\mathbf{y})}{\partial \beta} \\ \frac{\partial l(\beta, \mathbf{u}|\mathbf{y})}{\partial \mathbf{u}} \end{bmatrix} = \begin{bmatrix} -\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ -\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} + \mathbf{G}^{-1}\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} + \mathbf{G}^{-1}\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

e então,

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Essas equações são conhecidas como as equações de Henderson e permitem obter soluções para os efeitos fixos $\hat{\beta}$ e a predição para os efeitos aleatórios $\hat{\mathbf{u}}$. As soluções das equações são dadas por:

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (7)$$

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})) = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) \quad (8)$$

em que,

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}$$

De acordo com Littell et al.(2006), se \mathbf{R} e \mathbf{G} forem conhecidas, $\hat{\beta}$ é denominado como o melhor estimador linear não viesado (BLUE) enquanto $\hat{\mathbf{u}}$ é conhecido como melhor preditor linear não viesado (BLUP).

2.2.2 Estimação dos Parâmetros das Matrizes de Covariância \mathbf{R} e \mathbf{G}

Segundo Pinheiros e Bates (2000), dentre os vários métodos de estimação para modelos lineares mistos, os mais comumente utilizados são os métodos da Máxima Verossimilhança (MV) e Máxima Verossimilhança Restrita (REML). De acordo com Littell et al. (2006), embora existam vários métodos para estimação dos parâmetros de covariâncias (Métodos dos Momentos, Método da Estimação Quadrática Não-viesada de Variância Mínima, Método III de Henderson, Máxima Verossimilhança, Máxima Verossimilhança Restrita e Pseudo-Verossimilhança), o método da Máxima Verossimilhança Restrita (MVR) é indiscutivelmente o mais importante dentre eles.

O método da máxima verossimilhança restrita, desenvolvido por Patterson e Thompson (1971), visa estimar componentes de variância, sendo o mais recomendado para dados desbalanceados. Além disso, este método considera a perda de graus de liberdade resultante da estimação dos efeitos fixos e as estimativas caem sempre dentro do espaço paramétrico. Desta forma, em melhoramento genético o método da máxima verossimilhança restrita evita erros de pequenas amostras associadas aos efeitos fixos e dados provenientes de populações submetidas à seleção (YAMAKI, 2006). Segundo Simonato (2009), o método da máxima verossimilhança restrita supõe a normalidade dos dados, é iterativo e fornece sempre estimativas não negativas dos componentes de variância.

De acordo com Freitas (2007), o logaritmo neperiano (\ln) da função de verossimilhança obtida pelo método da máxima verossimilhança restrita é:

$$l(\mathbf{G}, \mathbf{R}) = -\frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{n-p}{2}\ln(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})' \\ \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}) - \frac{n-p}{2} \left[1 + \ln \frac{2\phi}{n-p} \right]$$

em que p é o posto da matriz \mathbf{X} .

Por meio do método da máxima verossimilhança restrita é possível obter as estimativas para os parâmetros de \mathbf{G} e \mathbf{R} . Essas estimativas são obtidas através de processos iterativos, pelo fato de não ser possível obter expressões analíticas para os parâmetros. Portanto, são utilizados algoritmos iterativos tais como Newton-Raphson, Escore de Fisher e o algoritmo Maximização das Esperanças (EM).

2.2.3 Métodos Iterativos

2.2.3.1 Algoritmo de Newton-Raphson

O algoritmo de Newton-Raphson (ou método de Newton) foi desenvolvido por Isaac Newton e Joseph Raphson e tem por objetivo estimar as raízes de uma função. Nesse método, é utilizada uma expansão em série de Taylor da função escore sobre a estimativa atual para obter a próxima estimativa. O algoritmo de Newton-Raphson pode ser utilizado para encontrar as estimativas de máxima verossimilhança.

Considere o logaritmo da função de verossimilhança, para o qual deseja-se encontrar o máximo em $\boldsymbol{\theta}$:

$$\frac{\partial l\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 0$$

A expansão de primeira ordem é dada por:

$$\frac{\partial l\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 0 = U(\boldsymbol{\theta}) \approx U(\boldsymbol{\theta}_{(0)}) + \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_{(0)})$$

Igualando a zero, tem-se:

$$U(\boldsymbol{\theta}_{(0)}) + \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_{(0)}) = 0$$

Isolando $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{(0)} - \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} U(\boldsymbol{\theta}_{(0)})$$

Essa equação pode ser utilizada iterativamente:

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_{(m)} - \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} U(\boldsymbol{\theta}_{(m)}) = \boldsymbol{\theta}_{(m)} + [J(\boldsymbol{\theta})]^{-1} U(\boldsymbol{\theta}_{(m)})$$

começando a partir de um valor inicial pré-especificado $\boldsymbol{\theta}_{(0)}$, sendo $J(\boldsymbol{\theta})$ a matriz de informação observada em $\boldsymbol{\theta}_m$.

2.2.3.2 Algoritmo Escore de Fisher

O algoritmo Escore de Fisher foi proposto por Patterson e Thompson (1971) com o objetivo de obter as estimativas através do método da máxima verossimilhança restrita. Este procedimento consiste em substituir a matriz de informação observada pela matriz

de informação esperada no algoritmo Newton-Raphson. Assim,

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_{(m)} - E \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} U(\boldsymbol{\theta}_{(m)}) = \boldsymbol{\theta}_{(m)} + [I(\boldsymbol{\theta})]^{-1} U(\boldsymbol{\theta}_{(m)})$$

em que $I(\boldsymbol{\theta})$ é a matriz de informação esperada em $\boldsymbol{\theta}_{(m)}$.

2.2.3.3 Algoritmo Average Information

O algoritmo Average Information (AI) foi proposto por Johnson e Thompson (1995) e Gilmour et al. (1995). Este método se baseia no uso de uma matriz de informação média, consistindo da soma de quadrados de termos mais facilmente calculáveis.

Visto que as matrizes de informação observada e esperada envolvem o cálculo do traço de produtos de grandes matrizes, sendo de difícil computação (pois envolvem a segunda derivada), os autores propuseram o uso da matriz de informação média, a qual contempla uma média das matrizes de informação observada e esperada.

O cálculo da matriz de informação média é muito mais simples do que o cálculo de qualquer umas das duas matrizes (informação observada e esperada) isoladamente. Isto porque quando é feita a média das derivadas de segunda ordem observadas e esperadas, o termo envolvendo traços de produtos da matriz inversa são cancelados, permanecendo uma expressão de simples computação (RESENDE et. al., 2012).

2.2.3.4 Algoritmo EM (Maximização das Esperanças)

O algoritmo EM foi proposto por Dempster et al. (1977). Este algoritmo é um procedimento iterativo que converge para o estimador de máxima verossimilhança. O algoritmo EM consiste em completar os dados, ou seja, substituir os valores faltantes por valores estimados, o que torna o processo de maximização mais simples.

Este algoritmo envolve dois passos:

Passo E (Esperança): consiste em calcular $E(l(\boldsymbol{\theta}; \mathbf{Y}) | \mathbf{Y})$.

Passo M (Maximização): são as atualizações do vetor de sequência $\boldsymbol{\theta}_{(m)}$ pela maximização mais simples.

No passo E é calculada a esperança condicional da função log-verossimilhança dado o vetor de observações $E(l(\boldsymbol{\theta}; \mathbf{Y}) | \mathbf{Y})$ na m -ésima interação dado $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{(m)}$.

O passo M produz uma nova estimativa dos parâmetros para cada interação obtidas no passo E, repetindo o procedimento até que essa estimativa tenha convergido.

2.3 Modelo Misto no Melhoramento Genético Animal

No modelo misto utilizado para a avaliação genética em programas de melhoramento genético animal, os valores genéticos são considerados como efeitos aleatórios enquanto os efeitos não-genéticos são inclusos como efeitos fixos. Nesta abordagem, considera-se ainda a estrutura genealógica da população sob estudo, possibilitando a predição mais acurada do mérito genético (inclusive de animais não mensurados), a inclusão do coeficiente de endogamia para todos os animais avaliados e a análise de dados envolvendo múltiplas gerações. O modelo misto animal é dado por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (9)$$

em que: o vetor \mathbf{y} contém os fenótipos (observações) dos animais para determinada característica; \mathbf{X} é uma matriz de incidência relativa aos efeitos fixos; $\boldsymbol{\beta}$ é o vetor de efeitos fixos, que geralmente inclui os efeitos não-genéticos (tais como fazenda, ordem de lactação, estágio da lactação); \mathbf{Z} é a matriz de incidência relativa aos efeitos aleatórios; \mathbf{u} é o vetor dos efeitos aleatórios (valor genético), com $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{A}\sigma_a^2)$ em que \mathbf{A} é uma matriz construída a partir do relacionamento entre os animais e σ_a^2 é a variância genética aditiva; $\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios não observáveis.

Neste modelo, os valores genéticos são obtidos com propriedades de melhor preditor linear não viesado (*best linear unbiased predictor*, BLUP) (HENDERSON, 1975).

2.4 Metodologias Estatísticas em Seleção Genômica

2.4.1 BLUP Genômico (GBLUP)

Recentemente, com a disponibilidade de informações genômicas acerca dos animais, como os polimorfismos de nucleotídeo único (*single nucleotide polymorphisms*, SNP), métodos de avaliação genética baseados no relacionamento genômico entre os indivíduos tem sido propostos. Dentre estes, destaca-se o método GBLUP.

GBLUP é um método estatístico que emprega os relacionamentos entre os indivíduos, estimados a partir de SNP, para predizer o mérito genético de um indivíduo (VANRADEN, 2008). Trata-se portanto de um modelo no qual a matriz de relacionamentos baseada no pedigree é substituída por uma matriz derivada dos marcadores genéticos, denominada de matriz genômica (de los CAMPOS et al., 2013). Sendo assim, a covariância entre

indivíduos fundamenta-se na similiaridade observada ao nível genômico em substituição à similiaridade esperada por meio do pedigree, de forma que predições mais acuradas do mérito genético sejam obtidas (CLARK; van der WERF, 2013).

Segundo VanRanden (2008), o GBLUP pode ser aplicado através do modelo (1), no entanto, substituindo-se a matriz \mathbf{A} por uma matriz \mathbf{G} de relacionamentos genômica, de forma que $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{G}\sigma_g^2)$, sendo σ_g^2 a variância genética aditiva. A incorporação da informação de marcadores genéticos em uma matriz de relacionamentos foi primeiramente sugerida por Nejati-Javaremi et al. (1997), e até recentemente, diversos métodos foram propostos para a construção da matriz genômica. A matriz genômica proposta em VanRanden (2008) é construída como:

$$\mathbf{G} = \frac{\mathbf{W}\mathbf{W}'}{2 \sum p_i(1 - p_i)} \quad (10)$$

em que: p é a frequência do segundo alelo (alelo com a menor frequência) no *locus* i , \mathbf{W} é a matriz resultante da subtração de \mathbf{P} de \mathbf{M} , sendo $\mathbf{P} = 2(p_i - 0,5)$ e \mathbf{M} a matriz de genótipos codificados como 0, -1 e 1 para o heterozigoto e homozigotos, respectivamente. A correção da frequência para o alelo de menor frequência obriga a soma dos coeficientes entre os animais ser igual à zero para cada marcador enquanto que a divisão por $2 \sum p_i(1 - p_i)$ dispõe \mathbf{G} na mesma escala que a matriz de relacionamentos \mathbf{A} .

2.4.2 *single-step* GBLUP

Misztal et al. (2009), Legarra et al. (2009), Aguilar et al. (2010) e Christensen e Lund (2010) propuseram uma mudança ao método GBLUP. A matriz de relacionamentos baseada no pedigree (\mathbf{A}) é combinada com a matriz genômica (\mathbf{G}) formando uma matriz única denominada de \mathbf{H} (LEGARRA et al., 2009). Esse método é baseado na propriedade de que, em um dado *locus*, a covariância da frequência alélica ($g = 0; 0,5; 1$) para o i -ésimo alelo de dois indivíduos j e k é descrita pela expressão $Cov(g_{ij}, g_{ik}) = \theta_{jk}p_i(1 - p_i)$ (COCKERHAM, 1969), com θ_{jk} sendo a co-ancestralidade dos dois indivíduos (LEGARRA; DUCROCQ, 2012). Assim, é possível que os genótipos não observados de um determinado indivíduo sejam modelados por meio da matriz de relacionamentos baseada no pedigree (GENGLER et al., 2007; CHRISTENSEN; LUND, 2010). Em \mathbf{H} , a informação genotípica se propaga nos relacionamentos dos indivíduos não genotipados por meio de uma regressão linear que, implicitamente, prediz os genótipos não observados como com-

binações lineares dos genótipos observados, com os coeficientes de regressão derivados dos relacionamentos oriundos do pedigree (de los CAMPOS et al., 2013).

Sendo assim, a inversa da matriz \mathbf{H} (\mathbf{H}^{-1}) (LEGARRA et al., 2009; CHRISTENSEN; LUND, 2010) será formada por:

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix} \quad (11)$$

na qual os sub-índices 1 e 2 representam os animais não genotipados e genotipados, respectivamente; e \mathbf{G}^{-1} é a inversa da matriz de relacionamentos genômica \mathbf{G} (VANRADEN, 2008). Por sua vez, as equações de modelos mistos baseadas em (1) e na matriz anterior serão dadas por:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda_g \mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

com λ_g sendo a razão entre a variância residual (σ_e^2) e a variância dos efeitos genéticos poligênicos (σ_g^2).

Objetivando obter simplificação computacional, Aguilar et al. (2010) propuseram como inversa de \mathbf{H} :

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix} \quad (12)$$

O método *single-step* GBLUP foi utilizado em estudos envolvendo seleção genômica em frangos (Chen et al., 2011), suínos (Forni et al., 2011) e bovinos de leite (Aguilar et al., 2010, 2011; Tsuruta et al., 2011).

2.4.3 Construção da Matrix H

Existem algumas formas de modificação da matriz \mathbf{H}^{-1} de maneira que o método *single-step* GBLUP apresente melhor desempenho quanto à acurácia e controle do viés de predição. Dentre estas, tem sido reportada a introdução de um fator de escala (ω) para a sub-matriz \mathbf{A}_{22}^{-1} (TSURUTA et al, 2011; LOURENCO et al., 2014). Outra estratégia é a contabilização da proporção (α) da variância poligênica não explicada pelos SNP (AGUILAR et al., 2010). Esta última é realizada por meio da ponderação da informação

genômica e poligênica na matriz de relacionamentos \mathbf{G} . Dessa forma, tem-se:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{*-1} - \omega \mathbf{A}_{22}^{-1} \end{bmatrix} \quad (13)$$

com $\mathbf{G}^* = (1 - \alpha)\mathbf{G} + \alpha\mathbf{A}_{22}$.

3 MATERIAIS E MÉTODOS

3.1 Animais e Dados

Foram utilizadas 62.740 mensurações mensais de características produtivas de 6.022 vacas da raça Holandesa, obtidas junto a quatro fazendas entre maio de 2012 e dezembro de 2016. Para a obtenção das mensurações, houve a colaboração da Clínica do Leite, pertencente ao Departamento de Zootecnia da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP), localizada na cidade de Piracicaba-SP. Três fazendas ficam no estado de São Paulo, caracterizado por temperaturas médias mensais variando de 9,1 a 29,4 °C e pluviosidade anual de aproximadamente 1500 mm. A quarta fazenda fica no estado do Paraná, com temperaturas médias mensais entre 8,2 e 27,6 °C e pluviosidade anual de 1,971 mm. Em todas as fazendas, as vacas são ordenhadas três vezes ao dia por um sistema automático de ordenha, sendo mantidas em free-stall e alimentadas por ração total. Esta ração é composta por silagem de milho, feno de capim, caroço de algodão, farelo de soja, casca de soja, farelo de milho, polpa cítrica, minerais e vitaminas (PETRINI et al., 2016).

As informações disponibilizadas pelas fazendas foram genealogia, data de parto, dias em lactação e ordem de lactação adicionalmente aos dados de desempenho individual, como produção de leite (PROD, em kg), teor (g/100g de leite) de sólidos totais (SOLIDS) e de ácidos graxos (saturados (SAT) e insaturados (INSAT)).

Os grupos de contemporâneos (GC) foram formados associando-se o ano de parto, estação de parto (abril a setembro - estação seca; outubro a março - estação das águas), rebanho e mês e ano de análise, conforme proposto em Petrini et al. (2016). Antes das análises foram eliminadas informações referentes a animais com nenhum dado de desempenho em determinada data de coleta, e sem informação de data de parto, ordem de lactação e/ou idade. Além disso, foram eliminados registros fora da amplitude aceitável

(mais ou menos três desvios-padrão), associados a dias em lactação inferior a cinco ou superior a 305, a idade maior que 3.025 dias (aproximadamente oito anos) e também a ordem de lactação igual ou maior a sete. Ao final, foram compostos 462 GC e o número de medidas tomadas na primeira à sexta lactação foram, respectivamente, 27.318, 18.827, 10.360, 4.505, 1.508 e 222.

3.2 Genótipos

Foram ainda utilizados genótipos de 1.152 vacas para polimorfismos de nucleotídeo único (*single nucleotide polymorphism*, SNP). Destas, 763 foram genotipadas para um painel de baixa densidade, com 6.909 SNP (Bovine LD BeadChip, Illumina, San Diego, CA) enquanto 389 vacas foram genotipadas com um painel de média densidade, formado por 47.000 SNP (GeneSeek Genomic Profiler Bovine 50k, Neogen Corporation, Lincoln, NE). Todos os genótipos foram imputados para 79.294 SNP utilizando o programa findhap.v90 (VanRaden et al., 2013). O material genético utilizado para a genotipagem foi extraído dos bulbos pilosos da vassoura da cauda e utilizado o kit NucleoSpin Tissue kit (Macherey-Nagel GmbH e Co. KG, Düren, Alemanha). Marcadores com proporção de genótipos perdidos maior do que 10% ($n = 7.199$ SNP), monomórficos ou com frequência alélica menor do que 2% ($n = 1.206$ SNP), localizados nos cromossomos sexuais ($n = 2.485$ SNP), e amostras com call rate menor do que 90% ($n = 50$ animais) foram excluídos. Portanto, 1.102 vacas genotipadas para 68.404 SNP foram mantidas para as análises.

3.3 Análise Genética

3.3.1 Estimação dos Componentes de Variância

Para estimar os componentes de variância e prever os efeitos aleatórios foi utilizado o seguinte modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \boldsymbol{\varepsilon} \quad (14)$$

em que \mathbf{y} é o vetor dos fenótipos de interesse; \mathbf{X} é a matriz de incidência dos efeitos fixos; $\boldsymbol{\beta}$ é o vetor de efeitos fixos (ordem de lactação, grupo de contemporâneas, covariável do efeito cúbico de dias em lactação); \mathbf{Z} e \mathbf{W} são as matrizes de incidência dos efeitos aleatórios; \mathbf{u} é o vetor dos efeitos genéticos aditivos, com $\mathbf{u} \sim N(\mathbf{0}, \mathbf{H}\sigma_g^2)$, em que \mathbf{H} é a

matriz de relacionamentos genômica combinada e σ_g^2 é a variância genética aditiva; \mathbf{p} é o vetor dos efeitos aleatórios de ambiente permanente e $\boldsymbol{\varepsilon}$ é o vetor dos efeitos residuais, com $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$.

A inversa da matriz \mathbf{H} seguiu a estrutura proposta por Aguilar et al., (2010) e Christensen e Lund (2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{*-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix} \quad (15)$$

com $\mathbf{G}^* = (1 - \alpha)\mathbf{G} + \alpha\mathbf{A}_{22}$.

A herdabilidade (h^2) de cada característica foi calculada como a razão entre a variância genética aditiva e a variância fenotípica (obtida como a soma da variância genética aditiva, permanente de ambiente e residual). Similarmente, a proporção da variância fenotípica devida ao efeito permanente de ambiente (c^2) foi calculada como a razão entre a variância do efeito permanente de ambiente e variância fenotípica.

As estimativas dos componentes de variância e seus respectivos erros-padrão foram obtidos utilizando o programa AIREMLF90 da família de programas BLUPF90 (MISZTAL et al., 2019), por meio do método de máxima verossimilhança restrita (REML) com o algoritmo AI (*average information*).

3.3.2 Predição dos Valores Genéticos

O modelo (14) foi também utilizado na predição dos valores genéticos, considerando-se os valores estimados dos componentes de variância sob o respectivo modelo. A acurácia (r) dos valores genéticos foi estimada por meio da expressão:

$$r = \sqrt{1 - \frac{EP_j^2}{\hat{\sigma}_a^2}} \quad (16)$$

em que EP_j é o erro de predição associado ao valor genético do j -ésimo animal e $\hat{\sigma}_a^2$ é a variância genética aditiva da característica (HAYES et al., 2009).

Além disso, a tendência genética foi estimada pela regressão linear dos valores genéticos em relação ao ano de nascimento. O coeficiente de regressão foi considerado o ganho anual de cada característica. A regressão linear foi desenvolvida pela equação:

$$y_{ij} = \mu + b(x_{ij} - \bar{x}) + e_{ij} \quad (17)$$

em que y_{ij} é o valor genético predito para o j -ésimo animal do i -ésimo ano de nascimento, μ é o intercepto, b é o coeficiente angular associado ao ano de nascimento, x_{ij} é o i -ésimo ano de nascimento do j -ésimo animal, \bar{x} é a média do ano de nascimento, e e_{ij} é o resíduo, NID $(0, \hat{\sigma}^2)$. O coeficiente angular estimado corresponde ao ganho genético anual da característica estudada. Foi utilizado o teste t Student sob a hipótese de nulidade (H_0) de b ser igual a zero a um nível de significância de 5%. Sendo (H_0) rejeitada, conclui-se que houve progresso genético para as características estudadas.

4 RESULTADOS E DISCUSSÃO

4.1 Estatísticas Descritivas

As estatísticas descritivas para as características de produção de leite (PROD), ácidos graxos saturados (SAT), ácidos graxos insaturados (INSAT) e teor de sólidos totais (SOLIDS) estão submetidas na Tabela 1.

Tabela 1: Número de observações (N), média, desvio-padrão (DP), coeficiente de variação (CV, %), valores mínimos (MIN) e máximos (MAX) obtidos para produção de leite (PROD), ácidos graxos saturados (SAT), insaturados (INSAT) e teor de sólidos totais (SOLIDS).

Característica	N	Média	DP	CV	MIN	MAX
PROD (kg/dia)	62.740	34,112	10,262	30,08	2,7	65,3
SAT (g/100 g de leite)	28.365	2,263	0,533	23,55	0,42	4,59
INSAT (g/100 g de leite)	28.364	1,004	0,289	29,01	0,01	2,16
SOLIDS (g/100 g de leite)	38.384	12,26	0,969	7,9	6,61	17,71

Pode-se observar que o valor médio de teor de sólidos totais (SOLIDS) foi de 12,26 g/100 g de leite com coeficiente de variação igual a 7,9%. Costa et al. (2019) reportaram médias para teor de lactose, gordura e proteína igual a 4,76, 4,03, 3,38 respectivamente. Bobbo et al. (2020) relataram médias iguais a 3,92, 3,43 e 4,76 para teor de gordura, proteína e lactose respectivamente. Somando as médias das três características obtém-se um valor igual a 12,17 para o primeiro estudo e 12,11 para o segundo estudo. Esses valores se aproximam da média encontrada no presente estudo para sólidos totais, o qual

é composto em maior parte por gordura, proteína e lactose.

A produção de leite apresenta um valor médio de 34,11 kg/dia com coeficiente de variação igual a 30,08%. Valores próximos foram encontrados por Sahin et al. (2012), Tiezzi et al. (2013) e Penasa et al. (2015), em estudos para bovinos da raça Holandesa. Os autores obtiveram médias de 35,8; 32,00 e 31,6 kg/dia, respectivamente.

O valor médio de ácidos graxos saturados foi de 2,26 g/100 g de leite e foi próximo aos valores relatados nos estudos de Penasa et al., (2015) e Bobbo et al., (2020). Esses autores em média obtiveram teor de ácidos graxos saturados no leite de 2,58 g/100 g de leite e 2,62 g/100 g de leite, respectivamente. Para ácidos graxos insaturados, o valor médio foi de 1,004 g/100 g de leite e também foram próximos aos relatados por Bastin et al., (2011) e Penasa et al., (2015). Eles obtiveram médias de 1,31 g/100 g de leite e 1,11 g/100 g de leite, respectivamente.

4.2 Parâmetros Genéticos

Na Tabela 2 são apresentadas as estimativas dos componentes de variância, coeficiente de herdabilidade (h^2) e proporção da variância fenotípica devido ao efeito permanente de ambiente (c^2). Coeficientes de herdabilidade moderados foram estimados para ácidos graxos saturados e teor de sólidos totais. Para produção de leite e ácidos graxos insaturados a herdabilidade foi baixa.

Tabela 2: Componentes de variância e coeficientes de herdabilidade* (erro-padrão) estimados sob para produção de leite (PROD), ácidos graxos saturados (SAT), insaturados (INSAT) e teor de sólidos totais (SOLIDS).

Característica	$\hat{\sigma}_a^2$	$\hat{\sigma}_{pe}^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_p^2$	h^2	c^2
PROD	11,505 (1,287)	19,585 (1,023)	45,810 (0,273)	76,899 (0,016)	0,149 (0,823)	0,255 (0,013)
SAT	0,058 (0,005)	0,028 (0,004)	0,153 (0,001)	0,239 (0,003)	0,245 (0,02)	0,117 (0,016)
INSAT	0,005 (0,001)	0,004 (0,001)	0,057 (0,0005)	0,066 (0,001)	0,07 (0,011)	0,06 (0,009)
SOLIDS	0,231 (0,018)	0,086 (0,012)	0,501 (0,004)	0,818 (0,01)	0,282 (0,019)	0,105 (0,015)

* $\hat{\sigma}_a^2$: variância aditiva, $\hat{\sigma}_{pe}^2$: variância devido ao efeito permanente de ambiente, $\hat{\sigma}_e^2$: variância residual, $\hat{\sigma}_p^2$: variância fenotípica, h^2 : herdabilidade, c^2 : proporção da variância fenotípica devido ao efeito permanente de ambiente.

Similarmente, Iung et al. (2019), Costa et al. (2019) e Tiezzi et al. (2013) reportaram herdabilidades baixas para produção de leite no dia do controle leiteiro, sendo iguais a 0,13, 0,14 e 0,10, respectivamente. No primeiro estudo foram utilizadas informações de 5.981 vacas da raça Holandesa de primeira a sexta lactação. No segundo, foram utilizados 150.633 registros de 10.893 vacas de primeira a quinta lactação. No terceiro, foi considerada uma população de 16.089 vacas de primeira a nona lactação. No entanto, vários estudos estimaram herdabilidades maiores que no presente estudo. Charton et al. (2018), Rosca et al. (2018) e Schopen et al. (2009) estimaram herdabilidade igual a 0,41, 0,38 e 0,41 respectivamente. Petrini et al. (2016) explicam que essa diferença pode ocorrer devido ao uso de somente vacas de primeiro parto (primíparas). Os autores relatam que herdabilidades mais altas tem sido estimadas a partir de dados coletados na primeira lactação. Os autores Schopen et al. (2009) e Rosca et al. (2018) realizaram as análises com base na produção total de leite (kg na lactação). A característica de produção de leite no dia do teste (kg/dia) está relacionada com a produção total de leite. Porém, este fato não assume que as duas características sejam iguais, o que pode explicar a diferença das herdabilidades encontradas nos estudos citados acima e no presente estudo.

Não foram encontrados trabalhos recentes sobre parâmetros genéticos para o teor de sólidos totais no leite em populações de bovinos da raça Holandesa. No entanto, considerando-se que esta característica corresponde ao conteúdo de gordura, lactose e proteína, torna-se possível a comparação com as estimativas de herdabilidade obtidas para estas. Schopen et al. (2009), utilizando dados de bovinos da raça Holandesa, com 1.940 vacas de primeira lactação, estimaram herdabilidades altas para teor de gordura, proteína e lactose, variando entre 0,50 e 0,66, indicando que essas características podem ser utilizadas como critérios de seleção em programas de melhoramento genético animal. Similarmente, em Miglior et. al (2007), considerando uma população com 5.022 vacas da raça Holandesa de primeira à terceira lactação, também verificaram herdabilidade alta para essas características, sendo de 0,55 para teor de gordura, 0,58 para teor de proteína e 0,48 para teor de lactose.

Mais recentemente, Tiezzi et. al (2013), Penasa et. al (2015), Petrini et. al (2016), Rosca et al. (2018) estimaram herdabilidades próximas à característica de sólidos totais obtidas no presente estudo, para teor de gordura, teor de proteína e lactose. Estes estudos indicam a possibilidade de melhoria na qualidade do leite através da seleção de reprodutores. No primeiro, foram utilizadas informações de 16.089 vacas de primeira à nona lactação, obtendo herdabilidade igual a 0,28 para proteína e 0,33 para lactose. No segundo, foi considerada uma população de 25.317 vacas Holandesas múltiparas, com amostras obtidas entre 5 e 365 dias em lactação, em que a herdabilidade foi igual a 0,27 para proteína. No terceiro estudo, as herdabilidades para proteína e lactose foram iguais a 0,35 e 0,31, respectivamente, considerando uma população de 4.203 vacas da raça Holandesa, de primeira à sexta lactação. O quarto estudo considerou 1.897 registros de 998 vacas da raça Holandesa, obtendo herdabilidade igual a 0,27 para gordura e 0,25 para proteína.

Penasa et al. (2015), Iung et al. (2019) e Bobbo et al. (2020) reportaram estimativas de herdabilidade próximas ao presente estudo para características de ácidos graxos saturados e insaturados. Em Penasa et al. (2015) os valores de herdabilidade foram 0,25 para ácidos graxos saturados e 0,07 para ácidos graxos insaturados. Iung et al. (2019) estimaram herdabilidade igual a 0,25 para ácidos graxos saturados e 0,08 para ácidos graxos insaturados. Já Bobbo et al. (2020), considerando um total de 115.581 registros de vacas da raça Holandesa na Itália estimaram herdabilidade igual a 0,20 para ácidos

graxos saturados.

4.3 Valores Genéticos Preditos e Acurácia

Acurácias de médias altas (maiores que 0,70) foram estimadas para os valores genéticos preditos para teor de sólidos totais e ácidos graxos saturados (Tabela 3, Figura 1) enquanto para produção de leite e teor de ácidos graxos insaturados, tais médias foram de magnitude moderada. Conforme pode-se observar, maiores valores de acurácia foram obtidos para características com maior herdabilidade. Valores altos de herdabilidade indicam que uma maior proporção da variação fenotípica é oriunda da ação genética. Com isto, tem-se o fenótipo como um bom preditor do mérito genético e, portanto, obtém-se maiores valores de acurácia de predição.

Tabela 3: Média, desvio-padrão, máximo e mínimo da acurácia dos valores genéticos para produção de leite (PROD), ácidos graxos saturados (SAT), insaturados (INSAT) e teor de sólidos totais (SOLIDS).

Característica	Média	Desvio-Padrão	Mínimo	Máximo
PROD	0,662	0,067	0,355	0,802
SAT	0,703	0,111	0,104	0,872
INSAT	0,563	0,12	0,012	0,777
SOLIDS	0,758	0,096	0,168	0,895

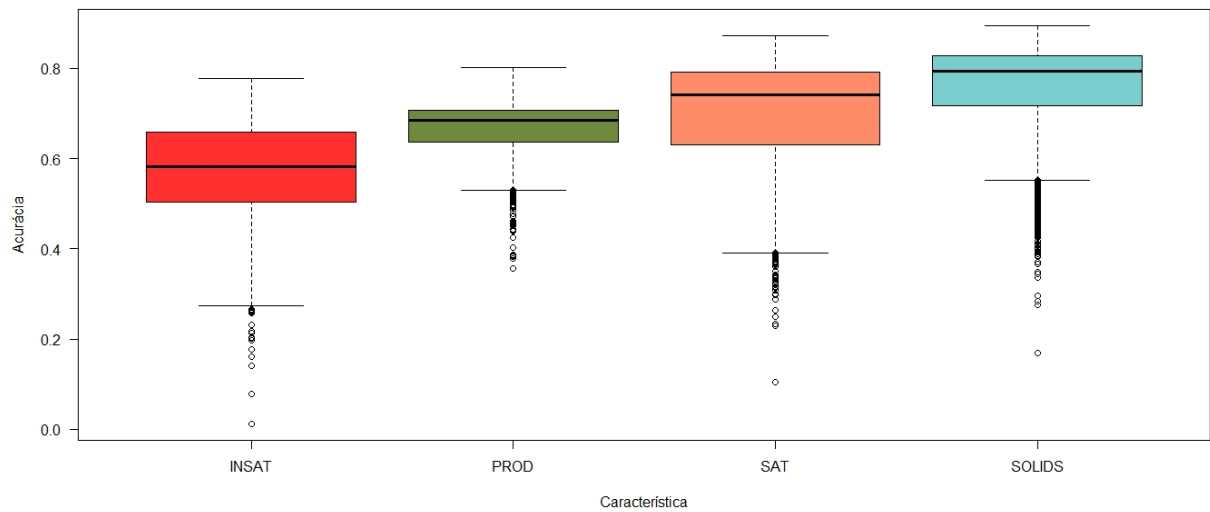


Figura 1: Acurácia dos valores genéticos preditos para ácidos graxos insaturados (INSAT), produção de leite (PROD), ácidos graxos saturados (SAT) e teor de sólidos totais (SOLIDS)

Maiores valores de acurácia também estão associados ao número de informações disponíveis. Por exemplo, o animal cujo valor genético foi predito com maior acurácia (0,895) para sólidos totais é uma vaca com 20 medidas de fenótipo e uma filha mensurada com 16 medidas de fenótipo, além de 176 meio-irmãs avaliadas. Por sua vez, o animal com menor acurácia (0,168) é uma vaca com cinco medidas de fenótipo e 58 meio-irmãs avaliadas.

4.4 Ganho Genético e Tendência Genética

Ganhos genéticos e tendências genéticas favoráveis foram verificados para produção de leite e teor de sólidos totais (Tabela 4; Figuras 4 e 5). Para ácidos graxos saturados e insaturados, os ganhos genéticos foram negativos ou nulos (Tabela 4; Figuras 2 e 3).

Tabela 4: Ganho genético (valor-p) predito para as características de produção de leite (PROD), ácidos graxos saturados (SAT), insaturados (INSAT) e teor de sólidos totais (SOLIDS).

Característica	Ganho Genético (valor-p)
PROD	0,0528 (< 0,0001)
SAT	-0,0004 (0,61)
INSAT	-0,0019 (< 0,0001)
SOLIDS	0,0073 (< 0,0001)

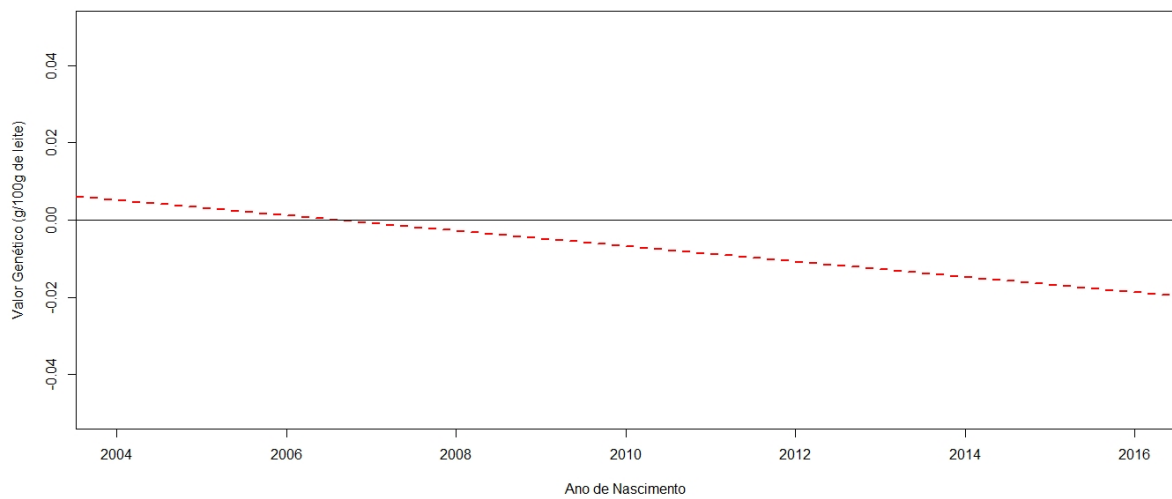


Figura 2: Tendência linear para teor de ácidos graxos insaturados (linha pontilhada vermelha).

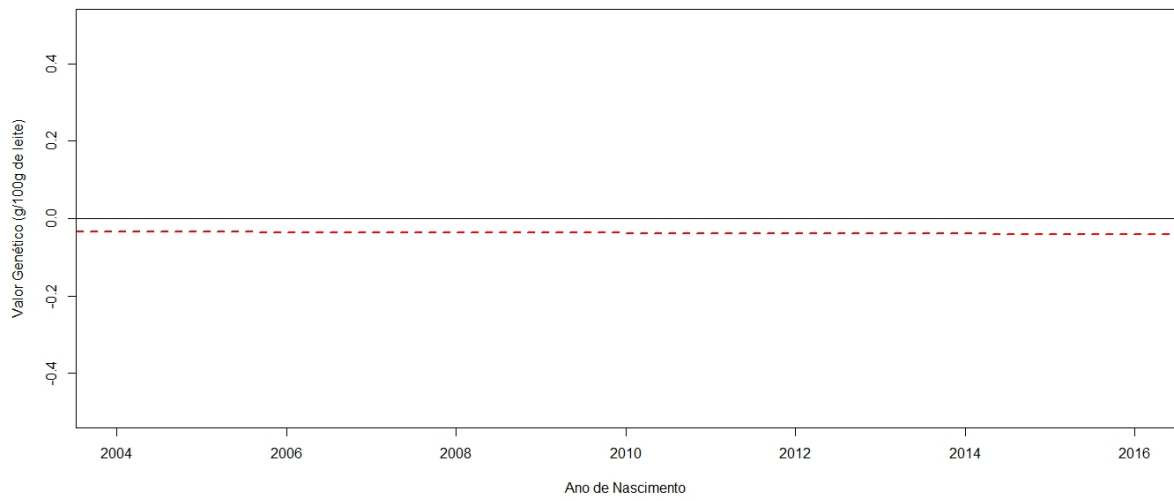


Figura 3: Tendência linear para teor de ácidos graxos saturados (linha pontilhada vermelha).

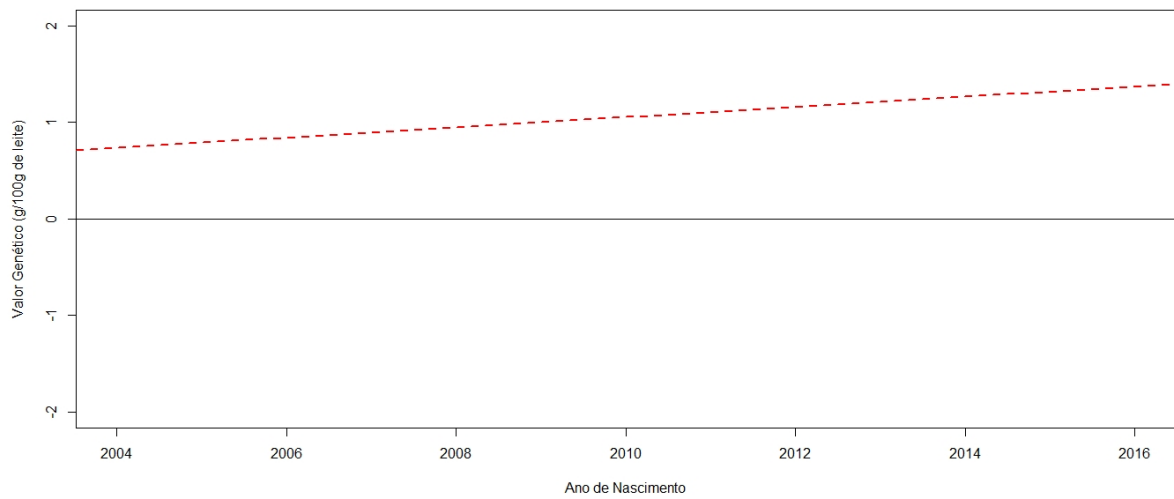


Figura 4: Tendência linear para produção de leite (linha pontilhada vermelha).

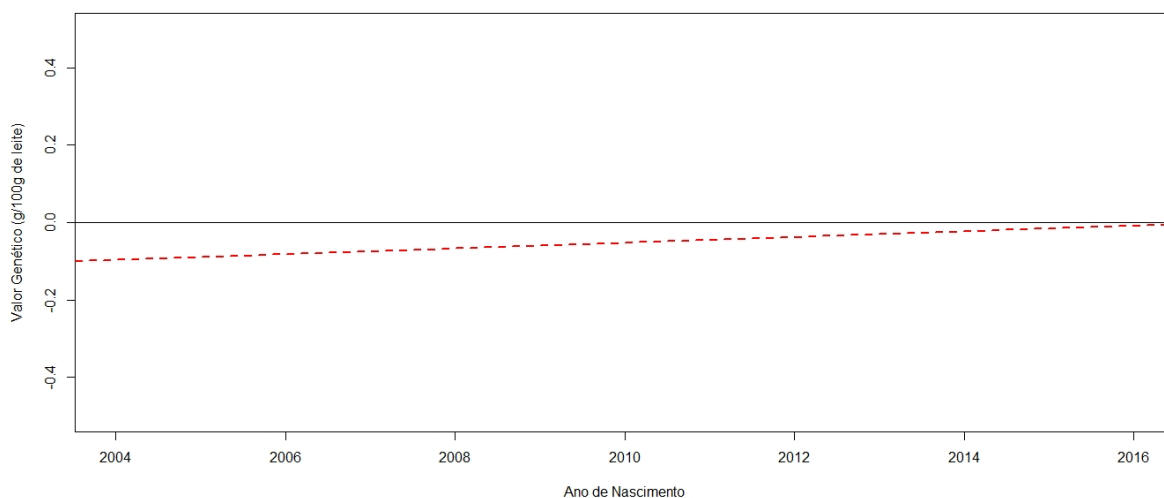


Figura 5: Tendência linear para teor de sólidos totais (linha pontilhada vermelha).

Para produção de leite e teor de sólidos totais a tendência genética foi positiva com aumento no valor genético principalmente a partir do ano de 2010. Esses resultados contribuem para uma possível mudança nos critérios de seleção, incorporando o uso de animais com maior valor genético para essas características.

Como pode-se observar pelas figuras 2 e 3, não foi verificado ganho genético para perfil de ácidos graxos.

5 CONCLUSÕES

Verificou-se que os componentes de variância estimados para teor de sólidos totais no leite indicam variabilidade genética, justificando a possibilidade de uso dessa característica como critério de seleção em programas de melhoramento animal visando-se ao aumento dos teores de gordura, proteína e lactose no leite, os quais, por sua vez, são utilizados na produção de derivados lácteos.

A tendência genética positiva para produção de leite e teor de sólidos totais contribui para o direcionamento da seleção, incorporando os animais com maior valor genético para essas características. De forma contrária, apesar da importância relacionada à qualidade do leite, o teor de ácidos graxos ainda não tem sido considerado na seleção de reprodutores. Isto possivelmente se deve à ausência de uma direção de seleção, dado que os ácidos graxos

estão associados a múltiplos fatores dentro do sistema de produção, desde à saúde da vaca até à qualidade nutricional do produto e suas características sensoriais.

A partir dos resultados obtidos, é possível concluir que as informações dos parâmetros e valores genéticos das características observadas neste estudo podem ser utilizadas para contribuir no delineamento de estratégias de seleção genômica, destacando o fato de que modelos lineares mistos são adequados na estimação dos componentes de variância das características estudadas.

REFERÊNCIAS BIBLIOGRÁFICAS

AGUILAR, I. et al. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, p. 743– 752, 2010.

AGUILAR, I. et al. Multiple trait genomic evaluation of conception rate in Holsteins. **Journal of Dairy Science**, v. 94, p. 2621-2624, 2011.

BASTIN, C.; GENGLER, N.; SOYEURT, H. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. **Journal of Dairy Science**, v. 94, n. 8, p. 4152-4163, 2011.

BOBBO, T.; PENASA, M.; CASSANDRO, M. Combining total and differential somatic cell count to better assess the association of udder health status with milk yield, composition and coagulation properties in cattle. **Italian Journal of Animal Science**, v. 19, n. 1, p. 697-703, 2020.

BOURDON, R. M. **Understanding animal breeding**. New Jersey: Prentice-Hall, 2000. 538 p.

CAMARINHA FILHO, J. A. **Modelos lineares mistos: estruturas de matrizes de variâncias e covariâncias e seleção de modelos**. 2016. 85p. Tese (Doutorado em Agronomia), Universidade de São Paulo, Piracicaba, 2002.

CARDOSO, V. et al. Breeding goals and economic values for pasture based milk production systems in the southeast region of Brazil. **Revista Brasileira de Zootecnia, Viçosa**, v. 33, n. 2, p. 320–327, 2004.

CHARTON, C. et al. Genetic parameters of milk production traits in response to a short once-daily milking period in crossbred Holstein× Normande dairy cows. **Journal of Dairy Science**, v. 101, n. 3, p. 2235-2247, 2018.

CHEN, C.Y. et al. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. **Journal of Animal Science**, v. 89, p. 23-28, 2011.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, p. 2-10, 2010.

CLARK, S. A.; van der WERF, J. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. IN: Gondro, C.; van der Werf, J.; Hayes, B. J. 23 (eds.). **Genome-wide association studies and genomic prediction, methods in molecular biology**. New York: Springer, 2013. chap. 13, p. 321-330.

COCKERHAM, C. C. Variance of gene frequencies. **Evolution**, p. 72-84, 1969.

COSTA, A. et al. Herdabilidade e repetibilidade da lactose do leite e suas relações com características tradicionais do leite, escore de células somáticas e ponto de congelamento em vacas Holandesas. **Animal**, v. 13, n. 5, pág. 909-916, 2019.

DE LOS CAMPOS. et al. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v. 193, p. 327-345, 2013.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 1, p. 1-22, 1977.

EENENNAAM, A. L. VAN.; YOUNG, A. E. Prevalence and impacts of genetically engineered feedstuffs on livestock populations. **Journal of Animal Science**, v. 92, p. 4.255-4.278, 2014.

FALCONER, D. S.; MACKAY, T. F. **Introduction to Quantitative Genetics**. England: Longman, 1996.

FERRAZ, J. B. S.; ELER, J. P. **Melhoramento genético para aumento de produtividade em gado de corte no Brasil: A história, o presente e o futuro.** [S.I.]: Bovinocultura de corte. v.2, 2010.

FERREIRA, T. A. **Características morfológicas e de tipo, divergência e avaliação genética de caprinos leiteiros registrados no Brasil de 1976 a 2009.** 2011. 71 p. Dissertação (Mestrado em Zootecnia), Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2011.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, v. 43, p. 1-8, 2011.

FREITAS, E. G. d. **Análise de dados longitudinais em experimentos com cana-de-açúcar.** 2007. 75 f. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP, 2007.

GARCIA, C. H. **Aplicação de equações de modelos mistos em testes clonais de *Eucalyptus spp.*** 2004. 86p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica)-Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2004.

GARCÍA-RUIZ, A.; COLE, J. B.; VANRADEN, P. M. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. **Proceedings of the National Academy of Sciences**, v. 113, n. 33, artigo E4928, 2016.

GENGLER, N.; MAYERES, P.; SZYDŁOWSKI, M. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. **Animal**, v. 1, n. 1, p. 21-28, 2007.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Average information REML: an ef-

ficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, p. 1440-1450, 1995.

GONZÁLES, F. H. D; NORO, D. Variações na composição do leite no subtropico brasileiro. **GONZÁLES, F. H. D et al. Qualidade do leite bovino, variações no trópico e subtropico. Passo Fundo, Brasil: LacVet**, p. 11-27, 2011.

HANUŠ, Oto et al. Role of fatty acids in milk fat and the influence of selected factors on their variability—a review. **Molecules**, v. 23, n. 7, p. 1636, 2018.

HARDING, F. **Milk Quality**. Dordrecht: Springer Science+Business Media , 1995. 178 p.

HAUG, A.; HØSTMARK, A. T.; HARSTAD, O. M. Bovine milk in human nutrition—a review. **Lipids in health and disease**, v. 6, n. 1, p. 1-16, 2007.

HAYES, B. J. et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. **Genetics Selection Evolution**, Les Ulis, v. 41, p. 51-60, 2009.

HAYES, B. J.; LEWIN, H. A.; GODDARD, M. E. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. **Trends in Genetics**, v. 29, n. 4., p. 206-214, 2013.

HENDERSON, C.R., KEMPTHORNE, O.; SEARLE, S.R.; VONKROSIGK, C.M. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v. 15, p. 192-218, 1959.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v. 31, p. 423-449, 1975.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA. **Produção da Pecuária Municipal**, Rio de Janeiro, v. 47, p. 1–8, 2019.

IUNG, L. H. S. et. al. Genome-wide association study for milk production traits in a Brazilian Holstein population. **Journal of Dairy Science**, v. 102, n. 6, p.5.305-5.314, 2019.

JOHNSON, D. L.; THOMPSON, R. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. **Journal of Dairy Science**, v. 78, n. 2, p. 449-456, 1995.

KRAG, K. et al. Genetic parameters for milk fatty acids in Danish Holstein cattle based on SNP markers using a Bayesian approach. **BMC Genetics**, v. 14, p. 79, 2013.

LEE, S. H. et al. Estimating missing heritability for disease from genome-wide association studies. **The American Journal of Human Genetics**, v.88, p. 294-305, 2011.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, p. 4656-4663, 2009.

LEGARRA, Andres; DUCROCQ, Vincent. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. **Journal of dairy Science**, v. 95, n. 8, p. 4629-4645, 2012.

LITTELL, R. C. et al. **SAS mixed models**. 2 ed. Cary: SAS Institute, 2006. 814 p.

LOURENCO, D. A. L.; MISZTAL, I.; TSURUTA, S. et al. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. **Journal of Dairy Science**, v. 97, p. 1742-1752, 2014.

LYNCH, M.; WALSH, B. **Genetics and Analysis of Quantitative Traits**. Sunderland: Sinauer Associates, 1998.

MACKAY, T. F. The genetic architecture of quantitative traits. **Annual review of Genetics**, v. 35, p. 303-339, 2001.

MIGLIOR, F. et al. Genetic analysis of milk urea nitrogen and lactose and their relationships with other production traits in Canadian Holstein cattle. **Journal of Dairy Science**, Champaign, v.90, p. 2468-2479, 2007.

MISZTAL, Ignacy; LEGARRA, Andres; AGUILAR, Ignacio. Procedimentos de computação para avaliação genética, incluindo informações fenotípicas, de linhagem completa e genômica. *Journal of Milk Science*, v. 92, n. 9, pág. 4648-4655, 2009.

MISZTAL, I. et al. Manual for BLUPF90 family of programs. Georgia, USA: Athens: University of Georgia, 2019.

MOURÃO, G. B.; FORMIGONI, I. B. Entendendo a seleção, seus conceitos e sua aplicação. In: TURCO, C. P. (Coord.). **Melhoramento genético ao alcance do produtor: bovinocultura de corte**. Bebedouro: Editora Scot Consultoria, cap. 2, p. 27-46., 2006.

NEJATI-JAVAREMI, A.; SMITH, C.; GIBSON, J. P. Effect of total allelic relationship on accuracy of evaluation and response to selection. **Journal of Animal Science**, v. 75, n. 7, p. 1738-1745, 1997.

PARODI, P. W. Milk fat in human nutrition. **Australian Journal of Dairy Technology**, v. 59, n. 1, p. 3, 2004.

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

PENASA, M. et al. Genetics of milk fatty acid groups predicted during routine data recording in Holstein dairy cattle. **Livestock Science**, Amsterdam, v.173, p. 9-13, 2015.

PEREIRA, C. P.; Milk nutritional composition and its role in human health. **Nutrition**, v. 30, p. 619-627, 2014.

PEREIRA, J. C. C. Melhoramento Genético aplicado à produção animal. **Editora FEPMVZ**, Belo Horizonte, 5a Ed., p. 62-95, 2008.

PETRINI, J. et al. Assessing the accuracy of prediction for milk fatty acids by using a small reference population of tropical Holstein cows. **Journal of Animal Breeding and Genetics**, v. 136, p. 456-463, 2019.

PETRINI, J. et al. Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. **Journal of Animal Breeding and Genetics**, v. 133, p. 384-395, 2016.

PINHEIRO, J. C.; BATES, D. M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. 528p.

RESENDE, M. D. V. et al. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa/Departamento de Estatística. 2012. 291 p.

ROSCA, N.; CREANGĂ, Ș.; IVANCIA, M.. Research Regarding Heritability and Repeatability Coefficients of some Milk Production Traits from Farms in Moldavian Region. **Scientific Papers Animal Science and Biotechnologies**, v. 51, n. 2, p. 157-161, 2018.

SAHIN, A. et al. Genetic and environmental parameters and trends for milk production of Holstein cattle in Turkey. **Italian Journal of Animal Science**, v. 11, n. 3, p. e44, 2012.

SIMM, G. **Genetic improvement of cattle and sheep**. Farming press, 1998.

SIMONATO, K. L. B. **Métodos da Máxima Verossimilhança e Máxima Verossimilhança Restrita na estimação de parâmetros do modelo de variabilidade espacial**. 104 p. Dissertação (Mestrado em Engenharia Agrícola), Universidade Estadual

do Oeste do Paraná, Cascavel, 2009.

SIMOPOULOS, A. P. An increase in the omega-6/omega-3 fatty acid ratio increases the risk for obesity. **Nutrients**, v. 8, n. 3, p. 128, 2016.

SCHOPEN, G. C. B. et al. Genetic parameters for major milk proteins in Dutch Holstein-Friesians. **Journal of Dairy Science Champaign**, v.92, p. 1182-1191, 2009.

TIEZZI, F. et al. Heritability and repeatability of milk coagulation properties predicted by mid-infrared spectroscopy during routine data recording, and their relationships with milk yield and quality traits. **Animal**, Cambridge, n. 7, v.10, p. 1592-1599, 2013.

TSURUTA, S.; MISZTAL, I., AGUILAR, I. et al. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. **Journal of Dairy Science**, v. 94, p. 4198-4204, 2011.

VAYEGO, S. **Uso de Modelos Mistos na Avaliação Genética de Linhagens de Matrizes de Frango de Corte**. 2007. 104 p. Tese (Tese (Doutorado em Melhoramento Genético Animal)) — Universidade Federal do Paraná, Curitiba, 2007.

VANRADEN, P.M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, p. 4414–4423, 2008.

VILELA, Duarte; DE RESENDE, J. C. Cenário para a produção de leite no Brasil na próxima década. In: **Embrapa Gado de Leite-Artigo em anais de congresso (ALICE)**. In: SIMPÓSIO SOBRE SUSTENTABILIDADE DA PECUÁRIA LEITEIRA NA REGIÃO SUL DO BRASIL, 6.; SEMINÁRIO DOS CENTROS MESORREGIONAIS DE EXCELÊNCIA EM TECNOLOGIA DO LEITE, 2., 2014, Maringá. Perspectivas para a produção de leite no Brasil: anais. Maringá: Universidade Estadual de Maringá, 2014., 2014.

WHITE, T.; HODGE, G. **Predicting breeding values with applications in forest**

tree improvement. Dordrecht: Kluwer Academic, 1989. 367 p.

WIGGANS, G. R.; VANRADEN, P. M.; COOPER, T. A. The genomic evaluation system in the United States: Past, present, future. **Journal of Dairy Science**, v. 94, p. 3202-3211, 2011.

YAMAKI, M. **Estimação de parâmetros genéticos de produção de leite e de gordura da raça pardo-suíça, utilizando metodologias frequentista e bayesiana.** 64 p. Dissertação (Mestrado em Zootecnia), Universidade Federal de Viçosa, Minas Gerais, 2006.

ZAITLEN, N.; KRAFT, P. Heritability in the genome-wide association era. **Human Genetics**, v. 131, p. 1655-1664, 2012.

ZHU, Z.; BAKSHI, A.; VINKHUYZEN, A. A. E. et al. Dominance genetic variation contributes little to the missing heritability for human complex traits. **The American Journal of Human Genetics**, v. 96, p. 377-385, 2015.

ZOCCAL, R. Leite de vacas felizes. In: RENTERO, N. (Ed.). **Anuario Leite.** EMBRAPA, 2020.