

UNIVERSIDADE FEDERAL DE MINAS GERAIS

**Modelos estatísticos para análise de
características longitudinais em bovinos de
corte sob coleta seletiva de dados**

Fernanda Albuquerque Merlo

Belo Horizonte

2017

Fernanda Albuquerque Merlo

**Modelos estatísticos para análise de
características longitudinais em bovinos de
corte sob coleta seletiva de dados**

Dissertação apresentada ao Programa de Pós-Graduação em Zootecnia da Escola de Veterinária da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de mestre em Zootecnia.

Área de concentração: Genética e Melhoramento Animal

Prof. Orientador: Dr. Fabio Luiz Buranelo Toral

Coorientador: Dr. Henrique Torres Ventura

Belo Horizonte

2017

*“A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original.”*

Albert Einstein

AGRADECIMENTOS

Agradeço a Deus, referencial absoluto em minha vida, pela força, luz e coragem a mim confiadas. Também O agradeço pela oportunidade de aprendizagem técnica, crescimento profissional e amadurecimento pessoal proporcionados por todos os desafios e intempéries ocorridos nesse período.

Aos meus pais, Maria Liliam e Clóvis, pelo exemplo de integridade e trabalho e por tornar essa caminhada possível. Aos meus irmãos, Maria Paula e Leonardo, pelo apoio, longas conversas e bons conselhos concedidos à caçulinha. Assim como nossos pais, vocês também são meu alicerce e meus melhores amigos. Amo todos vocês.

Aos meus sobrinhos tão queridos; João Pedro, Julia, Helena e Bentinho; pela inocência, alegria e brincadeiras. Vocês são minha paz interior e minha fonte de inspiração.

Ao meu doutor, Rauno Álvaro, pelo amor, compreensão com as minhas crises de mestrandia e por viver minhas escolhas como se fossem suas.

A todos amigos da ABCZ que me apoiaram desde quando se tornar mestre era apenas um sonho, em especial aos meus grandes amigos, João Gilberto, Rivaldo Borges, Gustavo Laterza e Fabiano Mendonça. À toda equipe do Pro-Genética, Luiz Josahkian, Gleida Marques, Lauro Fraga, Rafael Resende, Edson Filho e Flaviana Domingos e funcionários do escritório de Belo Horizonte pela cooperação e incentivo.

Ao meu orientador, professor Fabio Luiz Buranelo Toral, exemplo de dedicação e entusiasmo, pelos grandes ensinamentos e amizade.

À família que se formou no Laboratório (LADA), Fernanda Raidan, Virgínia Ribeiro, Larissa Kretli, Luiza Abreu, Gabriela Gouveia e Suzana Maruch por serem mães, psicólogas, amigas e professoras. Vocês foram anjos em minha vida e sem vocês eu jamais teria conseguido.

Ao professor José Aurélio Garcia Bergmann, à Dra. Fernanda Raidan e ao Dr. Henrique Torres Ventura pelas contribuições para execução e conclusão da pesquisa.

A todos que, de alguma forma, contribuíram para a realização desse projeto.

SUMÁRIO

| | |
|---|----|
| LISTA DE TABELAS | 5 |
| LISTA DE FIGURAS | 6 |
| RESUMO | 8 |
| ABSTRACT | 9 |
| 1. INTRODUÇÃO GERAL | 11 |
| 2. REVISÃO DE LITERATURA | 13 |
| 2.1. Metodologias estatísticas para análises genéticas | 13 |
| 2.2. Modelos estatísticos aplicados ao melhoramento animal | 15 |
| 2.3. Seleção..... | 19 |
| 2.3.1. Seleção e metodologias estatísticas | 19 |
| 2.3.2. Seleção e modelos estatísticos | 22 |
| 3.0. STATISTICAL MODELS FOR THE ANALYSIS OF LONGITUDINAL TRAITS IN BEEF CATTLE UNDER SEQUENTIAL SELECTIVE RECORDING..... | 26 |
| 3.1. Introduction | 27 |
| 3.2. Material and methods | 28 |
| 3.3. Results | 35 |
| 3.4. Discussion..... | 45 |
| 3.5. Conclusions | 50 |
| 4.0. IMPLICATIONS | 51 |
| 5.0. ACKNOWLEDGMENTS | 51 |
| 6.0. REFERÊNCIAS | 52 |

LISTA DE TABELAS

| | |
|--|----|
| Table 1. Summary statistics in different databases ¹ for body weight (BW) traits ² of Nellore young bulls in the performance test..... | 30 |
| Table 2. Posterior means (and highest posterior density interval with 95% of samples) of the genetic, residual and phenotypic correlations, in complete database (DB100), in the multi-trait model (above diagonal) and random regression model with linear spline polynomials (below diagonal) | 42 |
| Table 3. Posterior means (and highest posterior density interval with 90% of samples) of the genetic, residual and phenotypic correlations, in database with records of 85% heavier animals after selective sampling (DB85) for the multi-trait model (above diagonal) and random regression model with linear spline polynomial (below diagonal) | 43 |
| Table 4. Posterior means (and highest posterior density interval with 90% of samples) of the genetic, residual and phenotypic correlations, in database with records of 70% heavier animals after selective sampling (DB70) for the multi-trait model (above diagonal) and and random regression model with linear spline polynomial (below diagonal) | 44 |

LISTA DE FIGURAS

Figure 1. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in complete database for the single trait model (STM), multi-trait model (MTM) and random regression model with linear spline polynomials (RRM), respectively. The bars represent the highest posterior density interval with 95% of samples.....36

Figure 2. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in the complete database (DB100) and the database with records of 85% heavier animals and 70% heavier animals after selective sampling (DB85 and DB70, respectively) for the single trait model (STM). The bars represent the highest posterior density interval with 95% of samples.....38

Figure 3. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in the complete database (DB100) and the database with records of 85% heavier animals and 70% heavier animals after selective sampling (DB85 and DB70, respectively) for the multi-trait model (MTM) and random regression model with linear spline polynomials (RRM), respectively. The bars represent the highest posterior density interval with 95% of samples.....40

Figure 4. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in the complete database (DB100) and the database with records of 85% of animals and 70% of animals after random sampling (RS85 and RS70, respectively) for the single trait model (STM). The bars represent the highest posterior density interval with 95% of samples.....41

RESUMO

O objetivo deste trabalho foi verificar qual modelo estatístico é o mais adequado para análise de características longitudinais em rebanhos que realizam coleta seletiva de dados. Foram utilizados os pesos aos 330, 385, 440, 495 e 550 dias de idade de machos Nelore. Os modelos unicaracterístico (STM), multicaracterístico (MTM) e de regressão aleatória com polinômios lineares do tipo spline (RRM) foram comparados. O processo de coleta seletiva foi simulado por meio de amostragens sequenciais dos 85 e 70% animais mais pesados ao longo do tempo. As distribuições a posteriori dos parâmetros genéticos foram obtidas por meio do amostrador de Gibbs, com os três modelos mencionados anteriormente. As variâncias para peso obtidas por análise dos bancos de dados com coleta seletiva foram menores que aquelas obtidas para o banco de dados completo quando o modelo STM foi utilizado. Por outro lado, não houve diferença significativa nas variâncias, herdabilidades e correlações para a característica peso em análises do banco de dados com coleta seletiva e com registros completos, quando os modelos MTM e RRM foram utilizados. Os modelos multicaracterístico e de regressão aleatória são adequados para análise de características longitudinais em rebanhos que realizam coleta seletiva de dados.

Palavras chaves: herdabilidade, multicaracterístico, regressão aleatória, seleção, unicaracterístico, variância

ABSTRACT

The aim of the current study was to verify which statistical model is the most suitable for analysis of longitudinal traits in herds that use selective data collection. Weights at 330, 385, 440, 495 and 550 days of age of Nellore males were used. The single-trait (STM), multi-trait (MTM) and random regression models with linear polynomial spline (RRM) models were compared. The selective collection was simulated through sequential samplings of the 85 and 70% heavier animals over time. The posteriori distributions of the genetic parameters were obtained by the Gibbs sampler, with the three models mentioned above. The variances for weights obtained for the selective collection file analysis were smaller than those obtained for the complete file when the STM model was used. On the other hand, there was no significant difference between variances, heritabilities and correlations for weight in all files evaluated, when the MTM and RRM models were used. Multi-trait and random regression models are suitable for the analysis of longitudinal traits in herds that perform selective data collection.

Key words: heritability, multivariate, random regression, selection, univariate, variance

1. INTRODUÇÃO GERAL

O rebanho bovino do Brasil é de aproximadamente 220 milhões de cabeças, mas a média da produtividade ainda pode ser considerada baixa em relação aquelas obtidas em países como Estados Unidos, Argentina e Austrália (USDA, 2016). Neste sentido, o melhoramento genético pode contribuir para o aumento na produtividade da bovinocultura brasileira. O sucesso de um programa de melhoramento genético depende da definição dos objetivos de seleção e, também, da correta identificação do mérito genético dos candidatos à seleção (Harris et al., 1984). O mérito genético dos candidatos à seleção pode ser predito por meio das avaliações genéticas e essas, por sua vez, dependem do conhecimento de parâmetros genéticos.

A coleta de dados a campo é fundamental para assegurar a confiabilidade dos parâmetros genéticos e é uma das etapas mais onerosas do programa de melhoramento genético. Custos relacionados à mão de obra, à estrutura física necessária para mensurações dos animais, ao estresse e perda de produtividade em função dos manejos devem ser contabilizados a fim de se definir as características e os animais que devem participar do processo de mensuração.

Portanto, a coleta seletiva de dados, que consiste em determinar quais animais do rebanho serão medidos e descartar os demais, pode se tornar uma estratégia fundamental para reduzir custos com mensurações e manutenção de animais menos produtivos no rebanho. Além disso, os recursos financeiros provenientes da venda de animais de descarte são fundamentais para a manutenção do fluxo de caixa da operação saudável.

O próprio processo de seleção artificial (Henderson, 1975; Pollak e Quaas, 1981; Gianola e Fernando, 1988; Long et al., 1991) e o uso de modelos estatísticos inadequados podem levar à estimação de parâmetros genéticos imprecisos e tomadas de decisão que podem

prejudicar a evolução genética do rebanho. Entretanto, trabalhos que abordam a adequação de modelos estatísticos para dados sob coleta seletiva são raros (Nobre et al., 2003a).

Nesta dissertação, serão apresentados os resultados provenientes do uso de modelos estatísticos unicaracterístico, multicaracterístico e de regressão aleatória para análises genéticas do peso de bovinos em diferentes idades em rebanhos que praticam coleta seletiva de dados.

2. REVISÃO DE LITERATURA

2.1. Metodologias estatísticas para análises genéticas

Rotineiramente, informações fenotípicas são utilizadas como critérios de seleção a fim de promover o progresso genético (Lush, 1931). O valor fenotípico de um indivíduo é definido pelo somatório das contribuições genéticas, ambientais e suas interações. O uso de métodos estatísticos com pressuposições adequadas que ajustem o maior número de fatores não genéticos, que influenciem a característica escolhida como critério de seleção e que estimem de maneira precisa o valor genético de um animal é indispensável para garantir a eficiência do processo de seleção (Yates, 1934). Dessa maneira, considera-se como um método adequado para estimação de parâmetros genéticos aquele que possui alta probabilidade de ranquear corretamente os indivíduos avaliados (Henderson, 1977).

Dentre os métodos estatísticos conhecidos, o BLUP (melhor preditor linear não viesado) é o mais utilizado. Proposto em 1949 por Henderson, ele foi originado por uma modificação da metodologia de quadrados mínimos, que era, até então, utilizada para estimar os parâmetros genéticos e gerar as avaliações genéticas dos indivíduos. O BLUP tem como pressuposição habitual considerar que os valores genéticos de indivíduos da população base possuem distribuição normal com média zero e que as variâncias genéticas e as correlações são conhecidas (Henderson, 1973). Uma importante propriedade desse método é fornecer

avaliações genéticas não viesadas, ou seja, a esperança do preditor é igual a esperança da variável desconhecida que será predita (Henderson, 1975).

Mesmo após o surgimento do BLUP, ainda ocorriam problemas computacionais em virtude da dificuldade de se inverter a matriz de variância do vetor de observações, cuja importância está relacionada à resolução das equações provenientes da metodologia dos quadrados mínimos. Desse modo, também em 1949, Henderson criou a metodologia dos modelos mistos que não exigia a inversão dessa matriz e fornecia o mesmo resultado do índice de seleção modificado pelo BLUP (Henderson, 1974). Esse método foi assim chamado por incluir ao modelo dos quadrados mínimos, que apresenta apenas efeitos fixos em suas equações, o efeito aleatório genético aditivo. Essa inclusão possibilitou a redução na variância do erro de predição e consequente aumento da acurácia de ranqueamento dos animais a partir dos valores genéticos preditos nessas avaliações genéticas.

A metodologia dos modelos mistos pode ser aplicada de diferentes maneiras, como, por exemplo, no modelo animal, modelo reprodutor, entre outros. O modelo animal foi proposto por Henderson e Quaas (1976) e possibilita a inclusão de todas as informações de parentesco disponíveis para a análise e consequente obtenção do valor genético para todos os indivíduos presentes no pedigree, inclusive para animais que não apresentam registros para a característica avaliada. Já no modelo reprodutor, considera-se apenas o parentesco entre reprodutores, e por esse motivo, são obtidos valores genéticos apenas para os reprodutores.

2.2. Modelos estatísticos aplicados ao melhoramento animal

Vários modelos estatísticos são utilizados para a realização de avaliações genéticas e muito se discute acerca de qual seria o melhor a ser aplicado em cada situação. Contudo, deve-se atentar à que se refere tal superioridade de um modelo em relação ao outro. Um modelo pode ser considerado superior por reduzir o erro de predição, pela simplicidade matemática e computacional, por redução ou eliminação do viés ou por possuir outras propriedades desejáveis (Sorensen e Kennedy, 1984a).

O modelo unicaracterístico apresenta relativa facilidade matemática e computacional, e, de forma geral, pode ser descrito como:

$$y_{ij} = u + GC_i + a_j + e_{ij},$$

onde: y_{ij} representa o valor fenotípico; u , a constante geral presente em todas as observações;

GC_i , o efeito do grupo contemporâneo i ; a_j , o valor genético do animal j e e_{ij} , o efeito residual.

As pressuposições assumidas para os vetores dos efeitos aleatório genético (a) e residual (e) ao utilizar o modelo unicaracterístico são: $a \sim N(0, A\sigma_a^2)$ e $e \sim N(0, I\sigma_e^2)$, respectivamente; sendo que: N representa a distribuição normal; A , a matriz de parentesco; σ_a^2 , a variância genética aditiva; I , a matriz identidade e σ_e^2 , a variância dos resíduos. Dessa maneira, a covariância e a correlação entre as características são desconsideradas por esse modelo pois cada característica é avaliada separadamente (Misztal et al., 1993).

Diferentemente do modelo unicaracterístico, o modelo multicaracterístico considera a correlação existente entre as características (Walter e Mao, 1985; Misztal et al., 1993) e permite a inclusão de várias características na análise para a formação dos índices de seleção (Valverde et al., 2007; Boligon et al., 2009). A importância da contribuição de várias características envolvidas na análise para a avaliação genética de um animal foi destacada por Schaeffer (1984), salientando que a inclusão de um maior número de informações e a relação entre as mesmas em análises multicaracterísticas contribui para resgatar maior proporção de variância genética aditiva, possibilitando o consequente aumento de herdabilidade.

O sistema de equações de modelos mistos multicaracterístico possui maior número de equações e de incógnitas em relação ao modelo unicaracterístico. Dessa maneira, ao se utilizar o modelo multicaracterístico, todas essas equações são utilizadas para predizer o valor genético de um animal para a característica de interesse aumentando a acurácia de predição.

O modelo multicaracterístico pode ser descrito de forma semelhante ao modelo unicaracterístico apresentado anteriormente. Entretanto, para o modelo multicaracterístico, o efeito genético é descrito por meio da matriz de variância e covariância genética e o efeito do erro pela matriz de variância e covariância do resíduo. Assim, as pressuposições assumidas para os efeitos aleatórios genéticos e residuais quando se utiliza o modelo multicaracterístico são:

$$\begin{bmatrix} a_1 \\ \sim \\ a_2 \\ \vdots \\ a_p \\ \sim \end{bmatrix} \sim N(0, A \otimes G) \quad \text{e} \quad \begin{bmatrix} e_1 \\ \sim \\ e_2 \\ \vdots \\ e_p \\ \sim \end{bmatrix} \sim N(0, I \otimes R), \quad \text{onde } G, \text{ corresponde à matriz}$$

$$\begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 a_2} & \cdots & \sigma_{a_1 a_p} \\ \sigma_{a_1 a_2} & \sigma_{a_2}^2 & \cdots & \sigma_{a_2 a_p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{a_1 a_p} & \sigma_{a_2 a_p} & \cdots & \sigma_{a_p}^2 \end{bmatrix} \quad \text{e } R, \text{ corresponde à matriz} \quad \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} & \cdots & \sigma_{e_1 e_p} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 & \cdots & \sigma_{e_2 e_p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{e_1 e_p} & \sigma_{e_2 e_p} & \cdots & \sigma_{e_p}^2 \end{bmatrix}; \quad a_p, \sim$$

corresponde o vetor de soluções dos efeitos genéticos aditivos direto para a característica p ; e_p , o vetor dos resíduos para a característica p .

Análises de algumas características longitudinais, como o peso corporal, utilizando modelos multicaracterísticos têm como desvantagem a necessidade de ajustamento dos dados para pontos definidos previamente como padrões, acarretando em descarte de muitos dados e a consequente redução da acurácia (Nobre et al., 2003a). Desta maneira, o modelo de regressão aleatória pode ser uma alternativa às análises convencionais que utilizam modelos multicaracterísticos para estimar componentes de variância e parâmetros genéticos, como, por exemplo, análises de peso realizadas em diferentes idades de bovinos de corte (Albuquerque e Meyer, 2001; Nobre et al., 2003a; Sakaguti et al., 2003; Meyer, 2005; Dias et al., 2006).

A regressão aleatória possibilita a obtenção de parâmetros genéticos em qualquer idade dentro do intervalo considerado (Nobre et al., 2003a, 2003b; Sousa Júnior et al., 2010). Esse modelo proporciona aumento na acurácia da avaliação em função da eliminação de pré-ajustes nos dados exigidos pelos modelos multicaracterísticos e à possibilidade de se trabalhar com todas as pesagens disponíveis com as covariâncias adequadas (Meyer, 2004).

A equação que representa o modelo de regressão aleatória via polinômio spline pode ser descrita, de forma geral, como:

$$y_{ijk} = GC_j + \sum_{h=1}^H \varphi_h(A_k)c_h + \sum_{h=1}^H \varphi_h(A_k)a_{i_h} + \sum_{h=1}^H \varphi_h(A_k)p_{i_h} + e_{ijk}$$

Em que: y_{ijk} , representa o fenótipo para a característica avaliada; H , o número de nós em cada modelo; $\varphi_h(A_k)$, o polinômio linear spline (PLS) h ; c_h , o coeficiente de regressão h para modelar a trajetória média; a_{i_h} , o coeficiente de regressão aleatório genético aditivo do animal i associado ao polinômio h ; p_{i_h} , o coeficiente de regressão aleatório para o efeito de ambiente

permanente do animal i associado ao polinômio h ; e_{ijk} , o erro associado a cada observação e os demais termos como descritos anteriormente.

As pressuposições assumidas para o efeito aleatório genético $\begin{bmatrix} a_1 \\ \sim \\ a_2 \\ \vdots \\ a_p \\ \sim \end{bmatrix}$, residual $\begin{bmatrix} e_1 \\ \sim \\ e_2 \\ \vdots \\ e_p \\ \sim \end{bmatrix}$ e

efeito de ambiente permanente $\begin{bmatrix} p_1 \\ \sim \\ p_2 \\ \vdots \\ p_p \\ \sim \end{bmatrix}$ quando se utiliza o modelo de regressão aleatória são:

$$\begin{bmatrix} a_1 \\ \sim \\ a_2 \\ \vdots \\ a_p \\ \sim \end{bmatrix} \sim N(\underline{0}, A \otimes K_a), \quad \begin{bmatrix} e_1 \\ \sim \\ e_2 \\ \vdots \\ e_p \\ \sim \end{bmatrix} \sim N(\underline{0}, I \otimes K_i), \quad \begin{bmatrix} p_1 \\ \sim \\ p_2 \\ \vdots \\ p_p \\ \sim \end{bmatrix} \sim N(\underline{0}, P \otimes K_p), \quad \text{onde } K_a, K_i, K_p$$

representam as matrizes de variâncias e covariâncias entre os coeficientes de regressão para os efeitos genético aditivo, residual e de ambiente permanente, respectivamente.

Dessa forma, as avaliações genéticas também podem ser realizadas por meio de modelos de regressão aleatória, utilizando diferentes tipos de polinômios como, por exemplo, o polinômio linear do tipo spline (Misztal, 2006). O uso do spline implica que as soluções encontradas para os coeficientes de regressão dos polinômios coincidam exatamente com as estimativas de componentes de variância para os pontos definidos pelos nós.

2.3. Seleção

2.3.1. Seleção e metodologias estatísticas

A relação entre os dados avaliados, a interação genótipo-ambiente e o processo de seleção podem dificultar a estimação de parâmetros genéticos. Assim, vários autores da área de melhoramento genético têm discutido a presença, a ausência e a quantificação do viés nas avaliações genéticas obtidas com o uso de diferentes métodos e modelos estatísticos em diversas condições da produção animal, principalmente em rebanhos em que a seleção é praticada (Long et al., 1991; Meyer et al., 1993; Nobre et al., 2003a, 2003b; Pedrosa et al., 2014).

A seleção artificial é capaz de provocar mudanças nas variâncias genéticas em função do desequilíbrio gamético existente em vários loci com efeitos aditivos (Bulmer, 1971), das covariâncias negativas geradas entre as frequências dos genes nos diferentes loci (Felsenstein, 1965), ou de ambos fatores simultaneamente (Sorensen e Kennedy, 1984a). Entretanto, Bulmer (1980) relata que, no âmbito da genética quantitativa, as características avaliadas são resultantes de grande número de processos fisiológicos, provenientes de grande número de proteínas e determinadas por grande um número de genes. Dessa forma, ele afirmou que o processo de seleção pode não ser capaz de influenciar as estimativas dos parâmetros genéticos dessas características quantitativas.

A seleção sequencial é caracterizada pela escolha de animais previamente selecionados para determinadas características. Em virtude desse processo, análises genéticas obtidas de rebanhos que praticam esse tipo seleção podem apresentar maior magnitude das diferenças observadas nas médias dos parâmetros genéticos em relação às análises de rebanhos que praticam outros tipos de seleção (Long et al., 1991; Schaeffer et al., 1997; Kaps et al., 1999). Dessa maneira, análises de rebanhos que praticam a seleção sequencial exigem tratamento estatístico diferente daqueles aplicados em análises de rebanhos que praticam apenas uma seleção durante todo ciclo produtivo (Schaeffer et al., 1997).

Na presença de seleção, o estimador dos quadrados mínimos não apresenta viés apenas quando os registros são devidamente ajustados para os efeitos fixos, a seleção ocorre dentro de uma geração e existe apenas um registro para cada candidato à seleção (Sorensen e Kennedy, 1984b). Já o estimador do modelo misto é considerado como não viesado e os valores genéticos apresentam a menor variância de predição do erro quando a seleção ocorre dentro dos níveis de efeito fixo e as variâncias dos efeitos aleatórios são conhecidas (Sorensen e Kennedy, 1984b). Nesse mesmo estudo, os autores verificaram que quando determinadas condições eram conhecidas, o modelo misto não requeria o uso de população controle para a adequada partição do fenótipo em componentes genéticos e ambientais.

Para que as equações dos modelos mistos que levam ao BLUP na ausência de seleção sejam aplicadas a dados gerados por um ou muitos ciclos de seleção, é necessário assumir que as variâncias dos efeitos aleatórios sejam corretamente especificadas e conhecidas, bem como que a distribuição normal multivariada continue após repetidas gerações de seleções e a seleção seja uma função linear dos registros baseada no valor fenotípico e que seja invariante aos efeitos fixos do modelo (Sorensen e Kennedy, 1984b).

Avaliando três modelos, quadrado mínimo com modelo reprodutor, quadrado mínimo com modelo animal e uma modificação do método de Henderson com modelo reprodutor na

presença ou ausência de seleção, Sorensen e Kennedy (1984a) demonstraram que, quando o acasalamento era ao acaso, o quadrado mínimo com modelo reprodutor foi o único método que apresentou viés, pois ignorava a endogamia e o parentesco não colateral entre animais. Já na presença de seleção, observou-se que o único método que não apresentou viés foi o quadrado mínimo com modelo animal. Além disso, os autores verificaram que a quantidade de viés provocada pela utilização de matriz de parentesco incompleta foi maior quando houve seleção do que quando ocorreu acasalamento ao acaso, sugerindo que para que haja eliminação do viés deve-se utilizar o modelo animal e a matriz de parentesco completa

Trabalhando com simulação, Van Tassel et al. (1995) avaliaram o quadrado médio do erro para os métodos de quadrados mínimos, máxima verossimilhança restrita e inferência bayesiana em três situações: seleção, amostragem aleatória e ausência de seleção. Não foram encontradas diferenças nas médias dos quadrados dos erros para os três métodos avaliados quando houve amostragem aleatória e ausência de seleção. Contudo, quando a seleção ocorreu, observou-se diferença no quadrado médio do erro nos métodos avaliados, levando os autores a concluir que as diferenças na acurácia dos métodos de estimação dos componentes de variância observadas ocorreram em função do efeito direto da seleção e não em virtude da mudança na estrutura da população. Sorensen e Kennedy (1984a) e Pieramati e Van Vleck (1993) também observaram diferença no quadrado médio do erro para características quantitativas na presença de seleção.

Outros resultados relacionados ao viés provocado pela seleção foram observados por Henderson (1949) que demonstrou que a metodologia dos quadrados mínimos leva à estimativas de parâmetros genéticos viesadas quando o efeito da vaca é considerado como fixo, por Lush e Shrode (1950) que explanaram acerca do viés nos parâmetros genéticos proveniente da estimação do fator de correção da idade e por Lush e Shrode (1950) e Henderson et al. (1959)

que demonstraram que o estimador dos quadrados mínimos para os parâmetros genéticos em modelos de seleção envolvendo performance a pasto é viesado.

2.3.2. Seleção e modelos estatísticos

As pressuposições relacionadas ao descarte de animais assumidas para o modelo multicaracterístico diferem daquelas assumidas para o unicaracterístico. No modelo multicaracterístico, essas pressuposições consideram o descarte como não aleatório e os melhores animais para determinada característica são escolhidos em detrimento dos demais (Bolígon et al., 2009). Os resultados encontrados por Pedrosa et al. (2014), avaliando modelos unicaracterístico, bicaracterístico e multicaracterístico com cinco características para peso ao nascimento, peso ao desmame, circunferência escrotal, ganho de peso e musculosidade, demonstraram que o modelo multicaracterístico foi capaz de eliminar o viés provocado pela seleção. Dessa forma, os autores concluíram que o modelo multicaracterístico que incluía as informações pré-desmame foi o mais recomendado para avaliações genéticas de características pós desmame por favorecer a redução do viés ocasionado pelo descarte de animais nas fases iniciais de criação.

O efeito da seleção sequencial pode ser amenizado pela inclusão de registros anteriores à seleção (Bulmer, 1971; Pollak et al., 1984; Henderson, 1986; Thompson e Meyer, 1986). A remoção do viés ocasionado pela seleção por intermédio da inclusão dessas informações pré seleção também foi confirmada por Meyer et al. (1993), Meyer (1995), Kaps et al. (1999), Mercadante et al. (2004), Bolígon et al. (2009) e Mashiloane et al. (2009) que compararam o modelo unicaracterístico com o multicaracterístico em que características mensuradas antes da

seleção foram incluídas. Esses autores observaram estimativas de parâmetros genéticos para características de crescimento pós desmame em bovinos de corte maiores nas análises multicaracterísticas em relação às unicaracterísticas. A correlação entre os valores genéticos para características de peso em análises realizadas com os modelos bi e multicaracterísticos demonstrou que a inclusão ou exclusão de informações pré-seleção foi capaz de promover alteração no ranqueamento dos animais avaliados (Mashiloane et al., 2009).

Pollak et al. (1984), utilizando informações de crescimento em touros, avaliaram dois modelos multicaracterísticos para peso que diferiam apenas pela inclusão do peso dos animais descartados ao desmame ou não e demonstraram que as análises que utilizaram apenas os dados selecionados foram viesadas pela seleção ao desmame. Dessa forma, os autores ressaltaram a importância da inclusão das informações pré-seleção na análise para o aumento da acurácia de predição.

Aumento na estimativa da herdabilidade para peso foi observado ao se utilizar análise bicaracterística em comparação a unicaracterística em ovinos da raça Santa Inês (Sousa et al., 1999). Da mesma forma, Sarmiento et al. (2006), comparando os modelos uni e multicaracterístico para análise do peso em um rebanho de ovinos da raça Santa Inês que não estava sob seleção, observaram estimativas de variância genética aditiva e de herdabilidades superiores quando o segundo modelo foi utilizado. Desse modo, foi demonstrado que a informação de uma característica contribuiu para a estimação das variâncias genéticas aditivas diretas da outra característica.

Os modelos multicaracterístico, de regressão aleatória com polinômio de Legendre e de regressão aleatória com polinômio spline foram avaliados para dados simulados por Bohmanova et al. (2005). Os dados divergiam entre si pelo número de registros apresentados para cada indivíduo e pela ausência ou presença de padronização dos dados. Não foram encontradas diferenças para as soluções quando os registros ocorreram nos pontos

padronizados. Para os registros que ocorreram fora das idades padrões, a acurácia foi mantida apenas para os modelos de regressão. Entretanto, avaliações genéticas menos acuradas quando a regressão aleatória foi utilizada em relação ao modelo multicaracterístico foram observadas por Meyer (1999) e Schaeffer et al. (1997).

O estudo de Nobre et al. (2003a) verificou o uso do modelo multicaracterístico e de regressão aleatória com polinômio de Legendre para avaliar dois arquivos de dados distintos, um contendo apenas dados completos e outro contendo dados incompletos para pesos em diferentes idades de bovinos. Ambos os modelos estudados apresentaram padrões de curvas de crescimento semelhantes para todos os arquivos avaliados. Todavia, as estimativas de variância genética e residual obtidas por meio da regressão aleatória foram superiores às aquelas obtidas por meio do modelo multicaracterístico no arquivo onde houve diminuição do contingente populacional, o que não foi observado quando os dados avaliados eram provenientes do arquivo completo. As estimativas de variâncias para o arquivo incompleto em relação ao completo foram superiores para os dois modelos avaliados. Já as estimativas de herdabilidade para peso foram diferentes entre os dois modelos avaliados em ambos arquivos de dados.

Diferenças nas estimativas de herdabilidade para peso também foram relatadas por Sousa Júnior et al. (2010). Porém esses autores, avaliando modelo bicaracterístico em relação à regressão aleatória, não encontraram diferenças nas estimativas de variância aditiva para os dois modelos como observado no trabalho de Nobre et al. (2003a).

Em outra pesquisa, Nobre et al. (2003b) verificaram que a correlação genética para diferença esperada na progênie (DEP) de peso em touros com número de filhos inferior a cinquenta foi distinta entre os modelos multicaracterístico e de regressão aleatória. Além disso, Nobre et al. (2003b) demonstraram que as variâncias para peso obtidas por meio do modelo de regressão aleatória eram mais flutuantes do que aquelas obtidas por meio do modelo multicaracterístico, especialmente em idades tardias. Dessa forma, os autores demonstraram

que a regressão aleatória pode ser imprecisa para estimar parâmetros genéticos em função de fatores como o tamanho do conjunto de dados, a presença de seleção, o tipo de modelo utilizado na análise e as metodologias aplicadas.

Em virtude do que foi discutido anteriormente, é relevante a necessidade do desenvolvimento de alternativas que viabilizem a obtenção de parâmetros genéticos estimados por meio de dados provenientes de populações que passaram por algum tipo de amostragem, como é o caso da grande maioria dos dados provenientes dos programas de melhoramento animal.

3.0. STATISTICAL MODELS FOR THE ANALYSIS OF LONGITUDINAL TRAITS IN BEEF CATTLE UNDER SEQUENTIAL SELECTIVE RECORDING

ABSTRACT. The comprehensive analyses of longitudinal traits under sequential selective recording could improve genetic parameters estimates and lead to more accurate selection decisions. The objective of this study was to evaluate statistical models for analyzing longitudinal traits under sequential selective recording. We estimated genetic parameters for 60,550 body weight records of 12,110 young Nellore bulls exploring three models: single trait (STM), multi-trait (MTM) and random regression model with linear spline polynomials (RRM). Four additional dataset were obtained from those whole data (DB100). Two of them through a sequential selective recording using 85% (DB85) and 70% of heaviest animals (DB70) and the other two dataset with the same number of records in each age however, random data collection of 85% (RS85) and 70% (RS70) of body weight records. In STM and MTM, the body weights were standardized at 330 (BW330), 385 (BW385), 440 (BW440), 495 (BW495) and 550 days of age (BW550). In RRM, the knots of linear splines polynomials were adjusted in the same ages that were standardized for STM and MTM and at 250 and 597 days of age. However, for RRM analysis the body weights were not standardized. In general, genetic parameters estimated by STM with DB100 were significantly different to those obtained with DB85 and DB70 and no significantly different to genetic parameters estimated with RS85 and RS70. Different percentages of selected individuals (DB85 and DB70) also provided different changes in genetic parameters. When MTM and RRM were employed no significant differences on genetic parameters estimates of body weight were observed across the DB100, DB85 and DB70. Thus,

both models, MTM and RRM, are adequate for genetic evaluation of the longitudinal traits under sequential selective data recording.

Keywords: heritability, multi-trait analysis, random regression, selection, single trait analysis

3.1. Introduction

The selection process can influence the estimates of variance components and the accuracy of breeding values (Long *et al.*, 1991; Schaeffer *et al.*, 1998; Kaps *et al.*, 1999). Thus, alternative models in which selection is practiced were previously reported by several authors (Nobre *et al.*, 2003a; Bohmanova *et al.*, 2005; Boligon *et al.*, 2009). However, the suitability of models for sequential selective recording was not addressed. Sequential selective recording involves the sequential measurement of a group of individuals for some longitudinal trait with the purpose of reducing costs related to labor and the management of the animals; therefore it is an alternative to commercial herds. This process could generate differences between the estimates of genetic parameters of herds that practice and do not practice it.

The single trait model (STM), multi-trait model (MTM) and random regression model with linear spline polynomials (RRM) are commonly employed to perform genetic evaluations. However, models that do not consider the correlation among random effects, such as the STM, may compromise the estimation accuracy of genetic parameters and breeding values (Misztal *et al.*, 1993). MTM and RRM enable the estimation of correlations among random effects and the inclusion of pre-selection records (Pollak *et al.*, 1984; Henderson, 1986; Thompson and Meyer, 1986 and Meyer, 1995). Thus, these models can reduce or eliminate the influence of sequential selection on genetic parameter estimates (Schaeffer *et al.*, 1998).

The MTM requires the standardization of data in analyses of longitudinal traits, such as weight, which can cause the loss of records obtained outside the range established as a reference (Meyer, 2004). Conversely, the RRM enables the inclusion of information without standardization and the estimation of breeding values for all animals within the evaluated period (Misztal, 2006).

The objective of this study was to evaluate statistical models, STM, MTM and RRM for analyzing longitudinal traits under sequential selective recording.

3.2. Material and methods

Original data and complete file

Data from young Nelore bulls that participated in performance tests on pasture of the Brazilian Zebu Breeders Association (Associação Brasileira dos Criadores de Zebu - ABCZ) were employed. The data were collected during a period of 294 days (70 days of adaptation and 224 days of testing), and the animals were weighed at the beginning of the adaptation period, at the end of the adaptation period and every 56 days until the end of the test.

For the implementation of the STM and MTM, the weights for the pre-established ages must be standardized (Meyer, 2004). The weights were standardized for the ages of 330 (BW330), 385 (BW385), 440 (BW440), 495 (BW495) and 550 days (BW550), which corresponded to the mean age at the performed weighing. The standardized weight at age X ($X = 330, 385, 440, 495$ and 550) was obtained by the following equation: $BW_X = ABW_X + ADG(X - AGE_X)$, where BW represents the standardized weight at age X ; ABW represents the actual weight obtained near age X ; ADG represents the mean daily gain for the period of 56 days prior to age X ; and AGE represents the actual age. For BW330, the ADG for standardization was obtained after the reference age.

The individual records that exceeded the range obtained from the mean weight in the test plus or minus 3.5 standard deviations and the records of all animals that participated in tests with less than 35 animals were excluded. After these omissions, the data file consisted of the weights of 12,110 animals, which participated in 226 performance tests that were performed between 2003 and 2012 in the North (Rondônia, Pará and Tocantins), Northeast (Bahia), Central-West (Goiás, Mato Grosso and Mato Grosso do Sul), Southeast (Espírito Santo, Minas Gerais and São Paulo) and South (Paraná) of Brazil.

The pedigree file was constructed from the genealogical records of the animals with validated data and their ancestors, which consisted of 70,212 animals in the numerator relationship matrix.

Sampling simulations

From the previously described data file, two files were constructed to simulate the selective recording processes. These files were formed by selective data recording of the heaviest animals at weighing. The percentages of individuals who were selected in each of the files were 85% (DB85) and 70% (DB70). If an animal did not satisfy the established criteria, the weights of the animal that were obtained at subsequent ages were excluded. The descriptive statistics of these three files are presented in Table 1.

Table 1. Summary statistics in different databases¹ for body weight (BW) traits² of Nellore young bulls in the performance test

| Database ¹ | Trait ² | n | Mean (kg) | SD (kg) | CV (%) |
|-----------------------|--------------------|--------|-----------|---------|--------|
| DB100 | BW330 | 12,110 | 224.9 | 37.7 | 16.76 |
| | BW385 | 12,110 | 245.5 | 43.0 | 17.52 |
| | BW440 | 12,110 | 279.7 | 46.4 | 16.59 |
| | BW495 | 12,110 | 317.8 | 50.0 | 15.73 |
| | BW550 | 12,110 | 347.9 | 50.4 | 14.49 |
| DB85 | BW330 | 12,110 | 224.9 | 37.7 | 16.76 |
| | BW385 | 10,406 | 251.7 | 41.7 | 16.57 |
| | BW440 | 8,955 | 288.9 | 45.1 | 15.61 |
| | BW495 | 7,707 | 331.5 | 47.8 | 14.42 |
| | BW550 | 6,664 | 366.0 | 47.5 | 12.98 |
| DB70 | BW330 | 12,110 | 224.9 | 37.7 | 16.76 |
| | BW385 | 8,583 | 256.0 | 41.3 | 16.13 |
| | BW440 | 6,120 | 296.4 | 45.1 | 15.22 |
| | BW495 | 4,380 | 342.2 | 47.7 | 13.94 |
| | BW550 | 3,156 | 379.3 | 47.7 | 12.58 |

¹ DB100 = database with 100% of records; DB85 and DB70 = database with records of 85% heavier animals and 70% heavier animals, respectively, after selective sampling;

² BW330, BW385, BW440, BW495 and BW550 = body weight at 330, 385, 440, 495 and 550 days of age; n = number of records; CV (%) = coefficient of variation.

Additional files were also formed by random sampling, with the same sampling percentages (RS85 and RS70). In these cases, the random sampling procedures were performed ten times according to each sampling percentage.

Statistical analyses

The data of the original file and the simulated files were analyzed using STM, MTM and RRM.

The general STM and MTM can be defined as $y_{ijk} = u + PT_j + b_j(A_k - \bar{A}_j) + a_i + e_{ijk}$

where y_{ijk} represents the observed value for the trait of the animal i in the weight gain test j at age k ; u denotes the general constant in each observation; PT_j represents the effect of the performance test j on the trait; b_j denotes the coefficient of linear regression of the effect of age on the trait nested in the performance test j ; A_k represents the age k ; \bar{A}_j denotes the mean age of the animals at the performance test j ; a_i denotes the genetic value of the animal i ; and e_{ijk} represents the error associated with each observation.

In the matrix notation, the STM can be defined as $\tilde{y} = X\tilde{\beta} + Z\tilde{a} + \tilde{e}$, where \tilde{y} represents the vector of observations; X denotes the incidence matrix of fixed effects; $\tilde{\beta}$ represents the solution vector for fixed effects; Z denotes the incidence matrix for random effects; \tilde{a} represents the solution vector for the individual genetic effects; and \tilde{e} denotes the residual effects vector.

The following assumptions were assumed for the effects included in the STM: a priori flat distributions, $\tilde{\beta}$ ($\tilde{\beta} \sim \text{constant}$); normal distributions for \tilde{a} ($\tilde{a} | A, \sigma_a^2 \sim N(0, A\sigma_a^2)$) and \tilde{e} ($\tilde{e} | \sigma_e^2 \sim N(0, I_{N_i}\sigma_e^2)$); and inverse chi-square for σ_a^2 ($\sigma_a^2 \sim \chi^{-2}(v_a, S_a^2)$) and σ_e^2 ($\sigma_e^2 \sim \chi^{-2}(v_e, S_e^2)$). In this case, A is the relationship matrix, I_{N_i} is the identity matrix for order N_i ($N_i = \text{number of observations}$); σ_a^2 is the additive genetic variance; σ_e^2 is the residual variance; and v_a , v_e , and S_a^2 , S_e^2 the hyperparameters of prior distributions.

In the matrix notation, the MTM can be represented by

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_5 \end{bmatrix} = \begin{bmatrix} X_1 & \Phi & \dots & \Phi \\ \Phi & X_2 & \dots & \Phi \\ \vdots & \vdots & \ddots & \vdots \\ \Phi & \Phi & \dots & X_5 \end{bmatrix} \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_5 \end{bmatrix} + \begin{bmatrix} Z_1 & \Phi & \dots & \Phi \\ \Phi & Z_2 & \dots & \Phi \\ \vdots & \vdots & \ddots & \vdots \\ \Phi & \Phi & \dots & Z_5 \end{bmatrix} \begin{bmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_5 \end{bmatrix} + \begin{bmatrix} \tilde{e}_1 \\ \tilde{e}_2 \\ \vdots \\ \tilde{e}_5 \end{bmatrix},$$

where y_h represents the vector with the observations of the trait h ($h_1 = \text{BW330}$, $h_2 = \text{BW385}$, $h_3 = \text{BW440}$, $h_4 = \text{BW495}$ and $h_5 = \text{BW550}$); X_h denotes the incidence matrix for fixed effects; β_h represents the vector of solutions for the fixed effects; Z_h denotes the incidence matrix of random effects; a_h represents the vector of solutions for the genetic value of each animal for the trait h ; and e_h denotes the residual effect vector associated with each observation of the trait h .

The following a priori distributions for the MTM were assumed: $\begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_5 \end{bmatrix}^t \sim$ constant, $\begin{bmatrix} a_1 & a_2 & \dots & a_5 \end{bmatrix}^t \left| A, G_0 \sim N\left(\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}^t, G_0 \otimes A\right)\right.$ and $\begin{bmatrix} e_1 & e_2 & \dots & e_5 \end{bmatrix}^t \left| R_0 \sim N\left(\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}^t, R_0 \otimes I_{N_i}\right)\right.$,

where G_0 represents the matrix of additive genetic (co)variances among the traits; \otimes denotes the direct product operator among the matrices; and R_0 represents the matrix of residual

(co)variances among the traits, such that $G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 a_2} & \dots & \sigma_{a_1 a_5} \\ \sigma_{a_1 a_2} & \sigma_{a_2}^2 & \dots & \sigma_{a_2 a_5} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{a_1 a_5} & \sigma_{a_2 a_5} & \dots & \sigma_{a_5}^2 \end{bmatrix}$ and

$R_0 = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} & \dots & \sigma_{e_1 e_5} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 & \dots & \sigma_{e_2 e_5} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{e_1 e_5} & \sigma_{e_2 e_5} & \dots & \sigma_{e_5}^2 \end{bmatrix}$. Inverse Wishart distributions were assumed for the covariance

matrices (5×5) $G_0 (G_0 \sim IW(\Sigma_a^2, n_a))$ and $R_0 (R_0 \sim IW(\Sigma_e^2, n_e))$, where Σ_a^2 , Σ_e^2 , n_a and n_e represent the hyperparameters of the inverted Wishart distributions.

The analyses using the RRM considered data without standardization. The RRM can

be represented by $y_{ijk} = PGP_j + \sum_{h=1}^7 \varphi_h(A_k) b_h + \sum_{h=1}^7 \varphi_h(A_k) a_{i_h} + \sum_{h=1}^7 \varphi_h(A_k) p_{i_h} + e_{ijk}$

where y_{ijk} represents the weight of the animal i in the performance test j at age k ; 7 denotes the number of knots; $\varphi_h(A_k)$ represents the linear spline polynomial h that refers to age k ; b_h denotes the coefficient of regression that is associated with the polynomial h that fits the effect of age k on weight; a_{i_h} represents the coefficient of the additive genetic random regression h for animal i ; p_{i_h} denotes the individual permanent environment random regression coefficient h for animal i ; and e_{ijk} represents the error associated with each observation.

The knots associated with the age effect were positioned at ages 250, 330, 385, 440, 495, 550 and 597 days. The inclusion of two extremes (250 and 597 days) was required to delineate the entire age range with available weights.

To implement linear polynomial splines, the age of the animal (k) was transformed into a covariate of the linear polynomial spline ($\varphi_h(A_k)$). Assuming h at the points $T_h (h = 1, \dots, n)$ and assuming that $T_h \leq t < T_{h+1}$, the (co)variables can be obtained in the knots h and $h + 1$ by the equations $\varphi_h(t) = \frac{T_{h+1} - t}{T_{h+1} - T_h}$ and $\varphi_{h+1}(t) = 1 - \frac{T_{h+1} - t}{T_{h+1} - T_h}$, respectively. For other values outside the ranges delimited by T_h and T_{h+1} , we have $\varphi_h(t) = 0$. Assuming that the

observed value for the standard age corresponds to the fitted node, $\varphi_h(T_k) = \begin{cases} 1 & \text{se } h = k \\ 0 & \text{se } h \neq k \end{cases}$

In matrix form, the RRM can be described as $\underset{\sim}{y}^* = X^* \underset{\sim}{\beta}^* + Z^* \underset{\sim}{a}^* + W^* \underset{\sim}{p}^* + \underset{\sim}{e}^*$,

where $\underset{\sim}{y}^*$ represents the vector of observations; X^* denotes the incidence matrix of the coefficients of the polynomials to model the fixed effects; $\underset{\sim}{\beta}^*$ represents the vector with the solutions for the fixed effects random regression coefficients; Z^* denotes the incidence matrix of the polynomial coefficients to model the direct additive genetic effects; $\underset{\sim}{a}^*$ represents the

vector with the solutions for the genetic regression coefficients; W^* denotes the incidence matrix of the permanent environment effect; p^* represents the permanent environment effect; and e^* denotes the vector of residual effects. Asterisks serve to differentiate elements of the random regression models from the other models.

For RRM, the following a priori assumptions were made: $\beta^* \sim \text{constante}$, $a^* | A, G_0^* \sim N(0, G_0^* \otimes A)$, $p^* | P_0^* \sim N(0, P_0^* \otimes I)$ and $e^* | R^* \sim N(0, R^*)$. Inverse Wishart distributions were assumed for the covariance matrices (7×7) $G_0^* (G_0^* \sim IW(\Sigma_a^{2*}, n_a^*))$, $P_0^* (P_0^* \sim IW(\Sigma_p^{2*}, n_p^*))$, where Σ_a^{2*} , Σ_p^{2*} , n_a^* and n_p^* represent the hyperparameters of these distributions. We considered the heterogeneity of the residual variance with $R = \text{diag}\{\sigma_e^2\}$, where σ_e^2 is the residual variance for BW330, BW385, BW440, BW495 and BW550. The residual variance corresponded to the sum of the permanent effect and the residual environment effect. Information about a posteriori complete conditional distributions are provided in Sorensen and Gianola (2002).

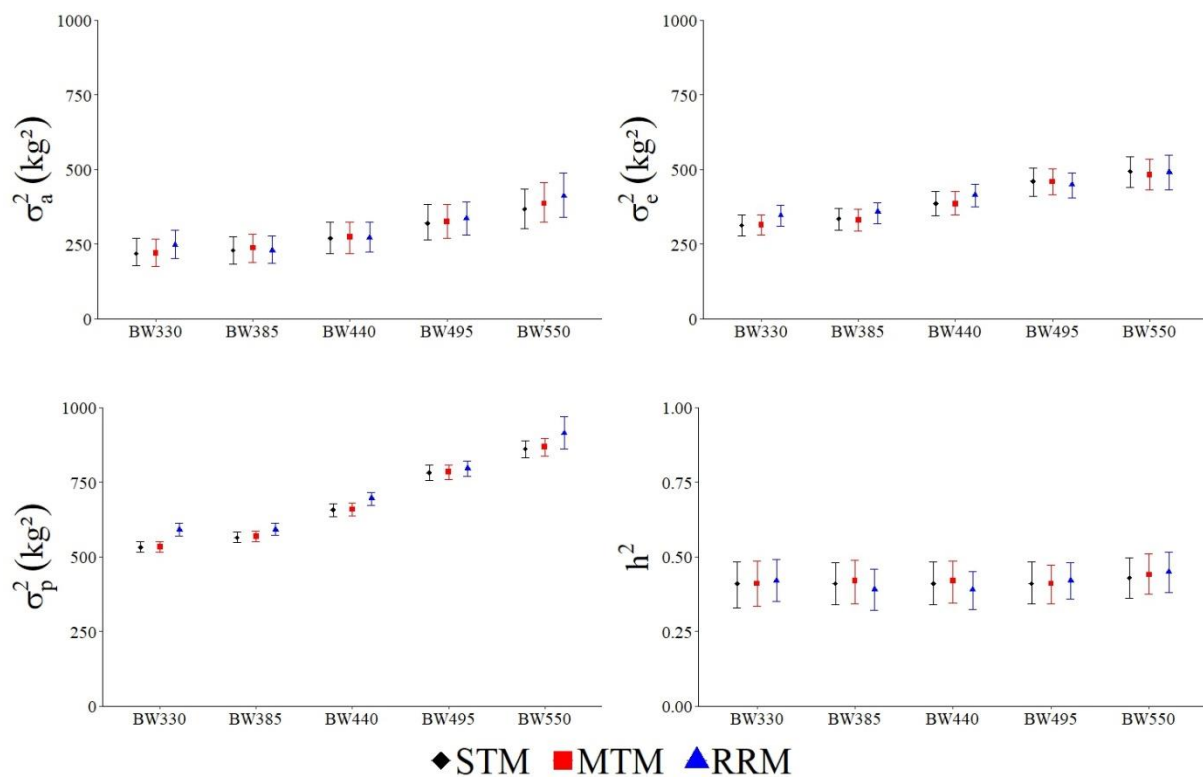
Samples of the complete conditional distributions of the covariance components of the models of interest were obtained by Gibbs sampler. The analyses were performed with GIBBS3F90 program (Misztal *et al.*, 2015). Chains of 410,000 samples were considered, 10,000 samples were discarded, and samples of covariance components values were obtained every 200 cycles. The chain size was defined in preliminary analyses according to the method of Raftery and Lewis (1992) and by visual inspection of the values sampled at each interaction, obtained by the BOA package (Smith, 2005) of the software R (R Development Core Team, 2015). The convergence diagnoses were performed using the criteria proposed by Geweke (1992) and Heidelberger and Welch (1983).

In case of the analyses with data from the files formed with random sampling, the samples of a posteriori distributions of each analysis were grouped to ensure that the estimates of the genetic parameters for the files RS85 and RS70 were obtained from 20,000 samples.

3.3. Results

The additive, residual and phenotypic variance and heritability estimated by STM, MTM and RRM for body weight measured in different ages using DB100 (Figure 1) were similar; since the highest density interval with 95% of samples of each genetic parameters estimated with the three models overlapped. Generally, the additive and residual variance increased between BW330 and the last two ages (BW440 and BW550) in the three studied models (Figure 1). Once the variances have increased proportionally, the heritability remained constant over the ages, with an approximate value of 0.40, regardless of the model used for its estimation (Figure 1). Still, an increase in the value of phenotypic variance estimate for the evaluated ages was observed. These estimates did not differ between BW330 and BW385 and started to increase significantly in BW440, up until BW550 (Figure 1).

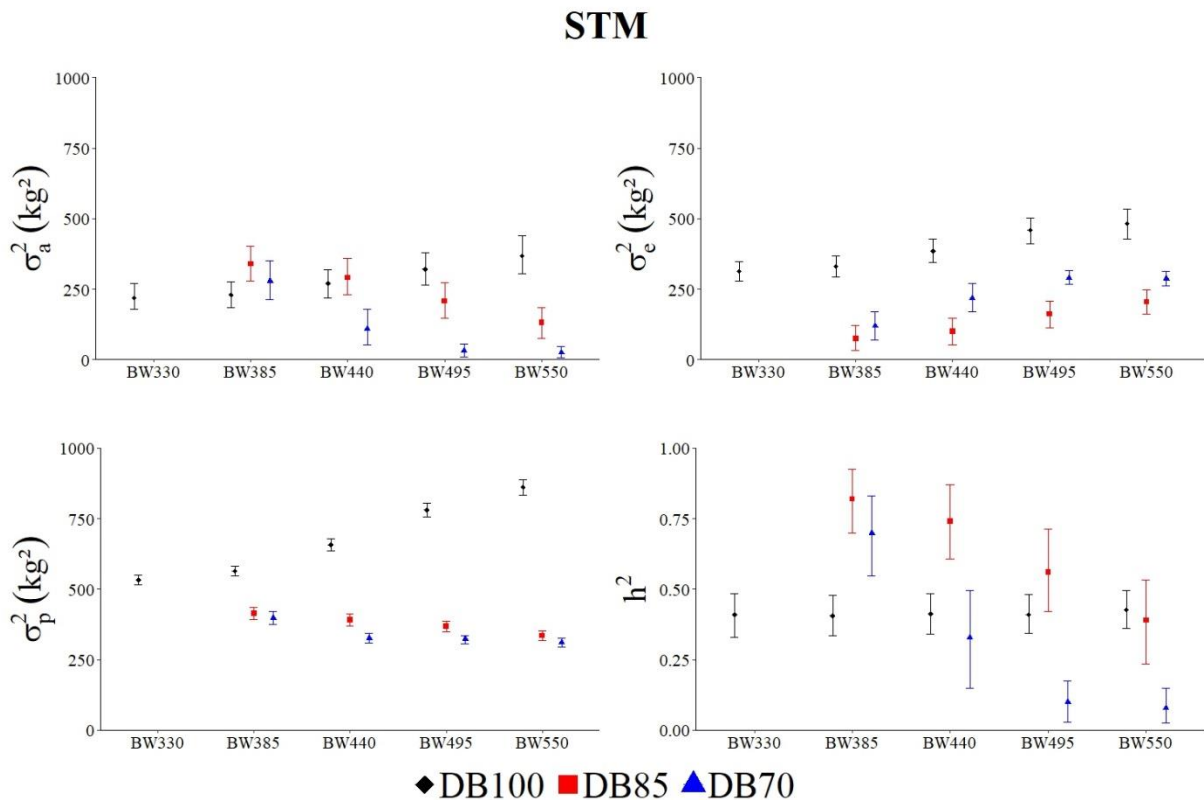
Figure 1. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in complete database for the single trait model (STM), multi-trait model (MTM) and random regression model with linear spline polynomials (RRM), respectively. The bars represent the highest posterior density interval with 95% of samples.



The behavior of genetic parameters when practicing selective recording was evaluated through the comparison between its estimates in DB100, DB85 and DB70. These comparisons were made in the three different models separately. Estimates obtained by each model and for each age with highest density interval with 95% of samples that overlapped were considered the same. The additive genetic variances estimates differed between the databases only when estimated by STM (Figure 2). At this point, the estimates for DB100 were higher than the

estimates for DB70 for BW495 and BW550. Still, in STM, the residual and phenotypic variances estimates of body weight with DB85 and DB70 were always lower than those obtained with DB100 (Figure 2). The heritabilities estimates obtained by STM differed between the databases. For BW385, the estimates obtained for DB85 and DB70 were higher than those obtained for DB100 (Figure 2). For BW440, this estimate differed only between DB100 and DB85, being higher in DB85. In contrast, for BW495 and BW550 the heritabilities estimates were lower for DB70 than for DB100 when estimated by STM (Figure 2). Comparing DB70 with DB85, there was decrease in genetic variance, phenotypic variance and heritability and increase in residual variance when the estimates were obtained by STM.

Figure 2. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in the complete database (DB100) and the database with records of 85% heavier animals and 70% heavier animals after selective sampling (DB85 and DB70, respectively) for the single trait model (STM). The bars represent the highest posterior density interval with 95% of samples.



The estimates of additive genetic and residual variances and heritabilities, when estimated by MTM and RRM did not differ between DB100, DB85 and DB70 over the ages (Figure 3). Phenotypic variance estimates, however, differed between the three databases at different ages. When estimated by MTM, phenotypic variances estimates were higher in DB100 than in DB85 and DB70 for BW550 (Figure 3). When estimated by RRM, two behaviors were

observed. For BW440, phenotypic variance estimates were lower in DB100 than in DB70 and, in contrast, for BW495 these estimates were higher in DB100 than in DB70 (Figure 3). The random reduction of the size of the database (DB100, RS85 and RS70) evaluated by STM did not cause changes in the estimates of the genetic parameters (Figure 4).

Figure 3. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in the complete database (DB100) and the database with records of 85% heavier animals and 70% heavier animals after selective sampling (DB85 and DB70, respectively) for the multi-trait model (MTM) and random regression model with linear spline polynomials (RRM), respectively. The bars represent the highest posterior density interval with 95% of samples.

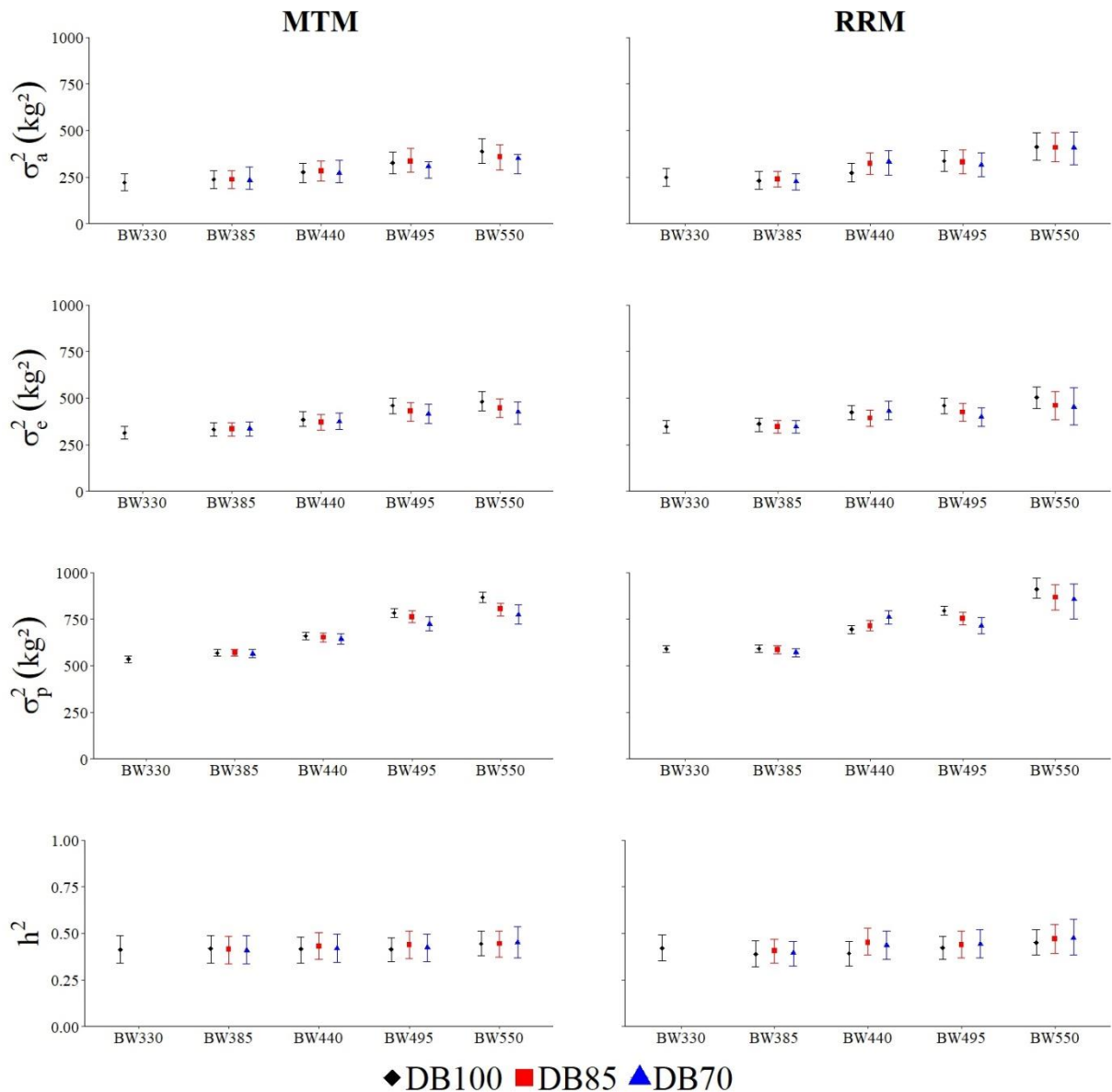
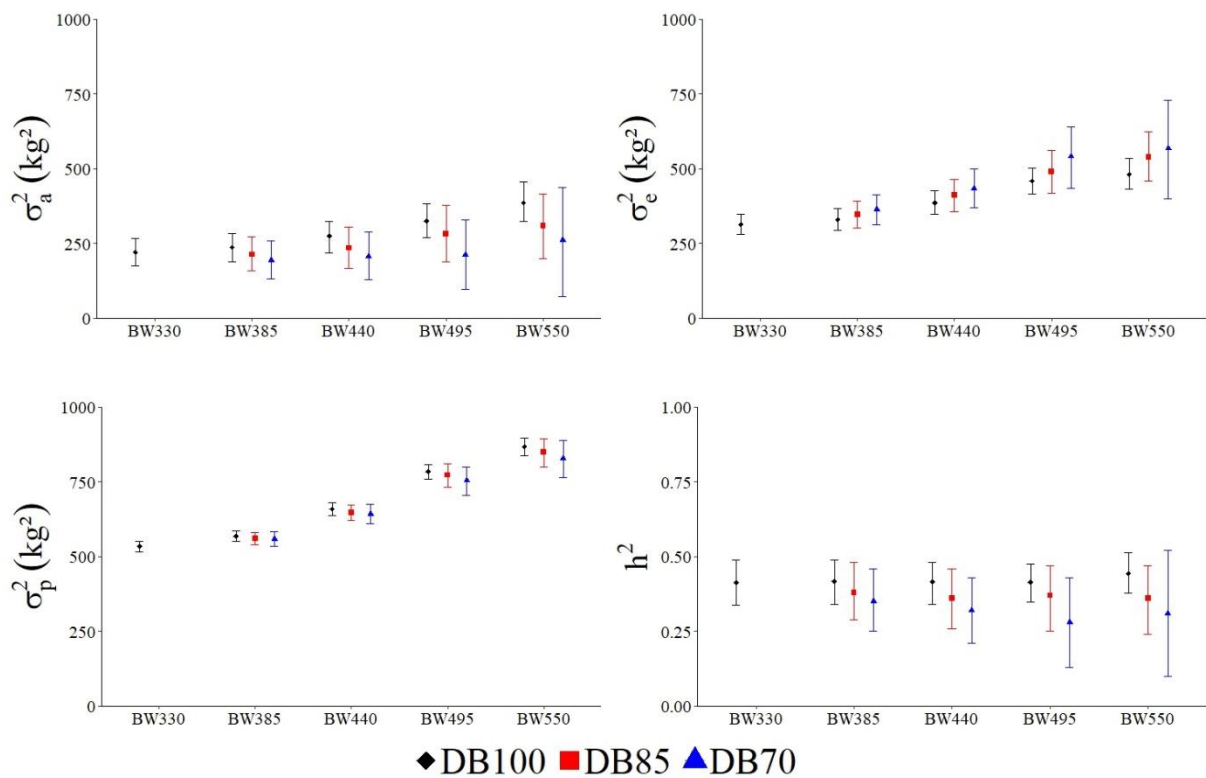


Figure 4. Posterior means of the genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability (h^2) for body weight at 330, 385, 440, 495 and 550 days (BW330, BW385, BW440, BW495 and BW550, respectively) in the complete database (DB100) and the database with records of 85% of animals and 70% of animals after random sampling (RS85 and RS70, respectively) for the single trait model (STM). The bars represent the highest posterior density interval with 95% of samples.



The correlations among the weights at different ages were high and were not influenced by the statistical model when estimated from DB100 (Table 2), DB85 (Table 3) and DB70 (Table 4). Furthermore, these estimates did not differ when comparing the same ages for DB100, DB85 and DB70. Generally, correlations between more distant ages are smaller than ages closer to each other (Tables 2, 3 and 4). The residual correlations were obtained from the

sum of the (co)variances of the permanent environment and residual variances when the RRM was used.

Table 2. Posterior means (and highest posterior density interval with 95% of samples) of the genetic, residual and phenotypic correlations, in complete database (DB100), in the multi-trait model (above diagonal) and random regression model with linear spline polynomials (below diagonal)

| Trait ¹ | BW330 | BW385 | BW440 | BW495 | BW550 |
|--------------------|-------------------------|----------------------|----------------------|----------------------|----------------------|
| | Genetic correlations | | | | |
| BW330 | - | 0.99 (0.98; 0.99) | 0.94 (0.92; 0.96) | 0.90 (0.87; 0.93) | 0.86 (0.82; 0.90) |
| BW385 | 0.98 (0.97; 0.99) | - | 0.98 (0.97; 0.99) | 0.95 (0.94; 0.97) | 0.92 (0.89; 0.94) |
| BW440 | 0.89 (0.85; 0.93) | 0.95 (0.93; 0.97) | - | 0.98 (0.98; 0.99) | 0.96 (0.95; 0.98) |
| BW495 | 0.86 (0.81; 0.90) | 0.93 (0.91; 0.96) | 0.97 (0.96; 0.98) | - | 0.99 (0.98; 0.99) |
| BW550 | 0.78 (0.72; 0.84) | 0.87 (0.83; 0.91) | 0.94 (0.91; 0.97) | 0.98 (0.96; 0.99) | - |
| | Residual correlations | | | | |
| BW330 | - | 0.89 (0.87; 0.90) | 0.78 (0.75; 0.80) | 0.73 (0.70; 0.76) | 0.71 (0.67; 0.74) |
| BW385 | 0.84 (0.82; 0.86) | - | 0.84 (0.82; 0.85) | 0.79 (0.77; 0.81) | 0.77 (0.74; 0.79) |
| BW440 | 0.71 (0.68; 0.74) | 0.81 (0.79; 0.83) | - | 0.84 (0.82; 0.85) | 0.82 (0.81; 0.84) |
| BW495 | 0.71 (0.67; 0.74) | 0.79 (0.77; 0.82) | 0.82 (0.80; 0.84) | - | 0.85 (0.83; 0.86) |
| BW550 | 0.68 (0.63; 0.74) | 0.76 (0.72; 0.81) | 0.80 (0.75; 0.85) | 0.87 (0.83; 0.92) | - |
| | Phenotypic correlations | | | | |
| BW330 | - | 0.93 (0.92; 0.93) | 0.84 (0.84; 0.85) | 0.80 (0.80; 0.81) | 0.77 (0.77; 0.78) |
| BW385 | 0.90 (0.89; 0.90) | - | 0.90 (0.89; 0.90) | 0.86 (0.85; 0.86) | 0.83 (0.82; 0.84) |
| BW440 | 0.78 (0.77; 0.79) | 0.87 (0.86; 0.87) | - | 0.90 (0.90; 0.90) | 0.88 (0.88; 0.89) |
| BW495 | 0.77 (0.76; 0.78) | 0.85 (0.84; 0.86) | 0.88 (0.88; 0.89) | - | 0.91 (0.90; 0.91) |
| BW550 | 0.72 (0.70; 0.75) | 0.81 (0.79; 0.83) | 0.86 (0.83; 0.88) | 0.92 (0.90; 0.94) | - |

Table 3. Posterior means (and highest posterior density interval with 90% of samples) of the genetic, residual and phenotypic correlations, in database with records of 85% heavier animals after selective sampling (DB85) for the multi-trait model (above diagonal) and random regression model with linear spline polynomial (below diagonal)

| Trait | BW330 | BW385 | BW440 | BW495 | BW550 |
|-------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Genetic correlations | | | | | |
| BW330 | - | 0.98 (0.98; 0.99) | 0.94 (0.92; 0.97) | 0.91 (0.88; 0.95) | 0.89 (0.83; 0.93) |
| BW385 | 0.97 (0.96; 0.98) | - | 0.98 (0.98; 0.99) | 0.96 (0.94; 0.98) | 0.94 (0.91; 0.96) |
| BW440 | 0.86 (0.81; 0.90) | 0.94 (0.92; 0.96) | - | 0.99 (0.98; 0.99) | 0.97 (0.95; 0.98) |
| BW495 | 0.87 (0.84; 0.92) | 0.95 (0.93; 0.97) | 0.97 (0.94; 0.99) | - | 0.99 (0.98; 0.99) |
| BW550 | 0.81 (0.75; 0.87) | 0.91 (0.88; 0.94) | 0.94 (0.91; 0.97) | 0.98 (0.96; 0.99) | - |
| Residual correlations | | | | | |
| BW330 | - | 0.89 (0.88; 0.90) | 0.78 (0.76; 0.81) | 0.74 (0.70; 0.77) | 0.71 (0.67; 0.74) |
| BW385 | 0.85 (0.84; 0.87) | - | 0.84 (0.83; 0.86) | 0.79 (0.76; 0.81) | 0.76 (0.73; 0.79) |
| BW440 | 0.73 (0.69; 0.76) | 0.82 (0.79; 0.84) | - | 0.84 (0.82; 0.86) | 0.82 (0.80; 0.85) |
| BW495 | 0.70 (0.67; 0.73) | 0.78 (0.75; 0.80) | 0.80 (0.77; 0.82) | - | 0.85 (0.84; 0.87) |
| BW550 | 0.64 (0.57; 0.70) | 0.71 (0.65; 0.77) | 0.75 (0.69; 0.81) | 0.82 (0.76; 0.88) | - |
| Phenotypic correlations | | | | | |
| BW330 | - | 0.93 (0.93; 0.93) | 0.85 (0.84; 0.86) | 0.81 (0.80; 0.82) | 0.78 (0.77; 0.79) |
| BW385 | 0.90 (0.90; 0.91) | - | 0.90 (0.90; 0.91) | 0.86 (0.86; 0.87) | 0.83 (0.82; 0.84) |
| BW440 | 0.78 (0.77; 0.80) | 0.87 (0.86; 0.88) | - | 0.90 (0.90; 0.91) | 0.89 (0.88; 0.89) |
| BW495 | 0.77 (0.76; 0.79) | 0.85 (0.84; 0.86) | 0.87 (0.86; 0.88) | - | 0.91 (0.91; 0.92) |
| BW550 | 0.71 (0.68; 0.74) | 0.80 (0.77; 0.83) | 0.84 (0.80; 0.87) | 0.89 (0.86; 0.92) | - |

Table 4. Posterior means (and highest posterior density interval with 90% of samples) of the genetic, residual and phenotypic correlations, in database with records of 70% heavier animals after selective sampling (DB70) for the multi-trait model (above diagonal) and random regression model with linear spline polynomial (below diagonal)

| Trait | BW330 | BW385 | BW440 | BW495 | BW550 |
|-------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Genetic correlations | | | | | |
| BW330 | - | 0.99 (0.98; 0.99) | 0.93 (0.90; 0.96) | 0.92 (0.87; 0.96) | 0.89 (0.84; 0.94) |
| BW385 | 0.98 (0.96; 0.99) | - | 0.97 (0.96; 0.99) | 0.96 (0.93; 0.98) | 0.93 (0.90; 0.97) |
| BW440 | 0.88 (0.83; 0.93) | 0.96 (0.93; 0.98) | - | 0.99 (0.98; 0.99) | 0.97 (0.95; 0.99) |
| BW495 | 0.89 (0.83; 0.94) | 0.94 (0.91; 0.97) | 0.97 (0.94; 0.98) | - | 0.99 (0.98; 0.99) |
| BW550 | 0.84 (0.77; 0.90) | 0.90 (0.85; 0.95) | 0.96 (0.93; 0.98) | 0.97 (0.95; 0.99) | - |
| Residual correlations | | | | | |
| BW330 | - | 0.89 (0.88; 0.90) | 0.79 (0.76; 0.82) | 0.73 (0.69; 0.76) | 0.70 (0.65; 0.74) |
| BW385 | 0.86 (0.84; 0.87) | - | 0.85 (0.84; 0.87) | 0.78 (0.75; 0.81) | 0.75 (0.72; 0.79) |
| BW440 | 0.73 (0.70; 0.77) | 0.83 (0.80; 0.85) | - | 0.83 (0.81; 0.85) | 0.79 (0.76; 0.82) |
| BW495 | 0.70 (0.66; 0.74) | 0.78 (0.76; 0.81) | 0.80 (0.77; 0.83) | - | 0.84 (0.82; 0.87) |
| BW550 | 0.64 (0.54; 0.73) | 0.73 (0.63; 0.82) | 0.76 (0.67; 0.86) | 0.79 (0.69; 0.88) | - |
| Phenotypic correlations | | | | | |
| BW330 | - | 0.93 (0.93; 0.93) | 0.85 (0.84; 0.86) | 0.81 (0.80; 0.82) | 0.78 (0.76; 0.79) |
| BW385 | 0.90 (0.90; 0.91) | - | 0.90 (0.90; 0.91) | 0.86 (0.85; 0.87) | 0.83 (0.82; 0.84) |
| BW440 | 0.80 (0.78; 0.81) | 0.88 (0.87; 0.89) | - | 0.90 (0.89; 0.91) | 0.87 (0.86; 0.88) |
| BW495 | 0.78 (0.76; 0.79) | 0.85 (0.84; 0.86) | 0.87 (0.86; 0.88) | - | 0.91 (0.90; 0.91) |
| BW550 | 0.72 (0.68; 0.78) | 0.80 (0.75; 0.85) | 0.85 (0.80; 0.90) | 0.87 (0.81; 0.92) | - |

3.4. Discussion

Sequential selection has an impact on genetic parameter estimates as can be seen when evaluating DB70 and DB85 using the STM model. However, MTM and RRM can help us deal with this situation.

The similarity between the additive genetic variance estimates and other genetic parameters obtained by the STM, MTM and RRM without selective recording (Figure 1) was also verified by Bohmanova *et al.* (2005). These authors performed a simulation study for genetic evaluation of growth traits in cattle and emphasized that the genetic parameters obtained by the MTM and RRM are equivalent when the records occur in standardized points. Similar to this study, the genetic parameters estimates obtained by STM, MTM and RRM did not differ. Thus, in situations where selective recording will not be practiced, the selection of the adequate model to process a genetic evaluation should be based on complementary aspects, such as an interest in the estimates of correlation between the evaluated traits and the facility of the analysis. For situations in which a researcher has an urgent need to know the components of the variance and genetic parameters from the analysis, the STM is recommended due to its mathematical and computational facility (Sarmiento *et al.*, 2006).

Due to the differences observed in the genetic parameters obtained by the STM in the DB85 and DB70 in relation to DB100 (Figure 2), the use of this model is not recommended for data analyses of selective recording herds. The differences in the genetic parameters estimates obtained by the STM are probably due to the exclusion of pre-selected records (Meyer, 1995; Kaps *et al.*, 1999; Mercadante *et al.*, 2004; Boligon *et al.*, 2009; Pedrosa *et al.*, 2014), the exclusion of covariances in the analysis (Walter and Mao, 1985) and the exclusion of other information about the herd, such as weight at different ages (Pollak *et al.*, 1984; Schaeffer *et al.*, 1998; Boligon *et al.*, 2009).

Differences in the estimates of additive genetic variance obtained by the STM in relation to the MTM and RRM in use of selective recording were observed (Figure 2 and 3). These results corroborate the findings of Mercadante *et al.* (2004) and Boligon *et al.* (2009), who obtained lower values of genetic additive variance for weight in analyses that employed the STM relative to the MTM.

Lower residual variance values obtained by the STM relative to the MTM were obtained in this study (Figure 2 and 3). Differences in the estimates of residual variance obtained in selective recording databases by the STM relative to the MTM were also obtained by Boligon *et al.* (2009) and Mashiloane *et al.* (2009). However, these authors attained higher values of residual variance obtained by the STM relative to the MTM. The selective data recording that was performed in this study generated a data structure for an analysis that differed from the data structure analyzed by these authors, which may explain the observed differences.

The selective recording, which is based on the phenotypic value, may favor the discarding of individuals from the same family, causing a reduction in the additive genetic variance. This fact, combined with the observed reduction of the residual variance estimates, caused a reduction of the phenotypic variance for DB85 and DB70 relative to DB100 when estimated by the STM for all evaluated ages (Figure 2). Differences in the phenotypic variance values were also observed when the STM was employed in relation to the MTM and RRM, which corroborates the results obtained by Boligon *et al.* (2009). However, these authors reported higher phenotypic variances when the STM was employed in relation to the MTM due to the significant increase in the residual variance for the STM, which differed from the residual variance observed in this study. The influence of selection on genetic parameters estimates, and consequently, the prediction of the genetic values of the animals, were demonstrated by Dickerson and Hazel (1944), Long *et al.* (1991), Schaeffer *et al.* (1998) and Kaps *et al.* (1999).

These authors emphasized the importance of the adequacy of the statistical models for the correct estimation and prediction of the genetic parameters.

The increase in the heritability obtained for BW385 observed in DB85 and DB70 in relation to DB100 by STM (Figure 2) can be explained by the reduction in the values of residual variance in relation to increase in the values of additive genetic variance. From this age, the difference between the residual variance and the additive genetic values decreased, and a consequent reduction in heritability was observed for the model in question.

The results obtained in this study corroborate those from Meyer (1995), Kaps *et al.* (1999), Mercadante *et al.* (2004), Sarmiento *et al.* (2006) and Pedrosa *et al.* (2014), who obtained higher estimates of genetic parameters for post-weaning growth traits in beef cattle using the MTM than the estimates obtained by the MTS with MTM. These results are attributed to the inclusion of traits measured prior to the selection in the analysis. Unlike the previously mentioned studies, this study also evaluated whether different percentages of selected individuals imply alterations in the differences among the genetic parameters estimates. The reduction in the additive, phenotypic genetic variances and heritability and the increase in the residual variance for DB70 in relation to DB85 when using the STM (Figure 2) indicate that the reduction in the percentage of selectively recorded data differently influences the estimates of the genetic parameters evaluated by this model.

The analysis of the data obtained by random sampling (RS85 and RS70) was performed to determine whether the differences in the genetic parameter estimates obtained using the STM were attributed to the selective recording or the reduction of the number of individuals in the data file. The results demonstrate that the reduction in the number of data from the file did not cause changes in the components of variance and heritability obtained via STM, therefore the difference found in the genetic parameters evaluated can be attributed to the sequential selection. Similarly, changes in the estimates of the genetic parameters caused by

restructures in the population, such as population reduction, were not verified by Van Tassel *et al.* (1995) and Nobre *et al.* (2003a). Thus, all differences in the genetic parameters among DB100, DB85 and DB70 were attributed to the selective recording when estimated by STM.

The effect of sequential selective recording on genetic parameters estimates can be mitigated by the inclusion of pre-selection records in the analysis (Bulmer, 1971; Pollak *et al.*, 1984; Henderson, 1986; Thompson and Meyer, 1986). In this study, the body weight records were measured at 330 days of age (BW330), at which time the selective recording process had not begun, which may have contributed to the suitability of the MTM for this type of evaluation. Multi-trait analyses of growth traits in bulls, which differed only by the inclusion or not of the body weight of animals discarded at weaning, were performed by Pollak *et al.* (1984). The authors verified the presence of bias in the evaluations in which no data of the discarded animals were employed due to the selection at weaning. The adequacy of the MTM for the analysis of the files with selective recording was favored due to the correlations between the traits in the analyses and the consideration of the missing values caused by the sequential selective recording (Meyer, 1995; Kaps *et al.*, 1999).

In genetic evaluations in which the correlation estimates are desired (Walter and Mao, 1985) and more than one selection criterion is simultaneously involved (Schaeffer, 1984), MTM and RRM are suitable alternatives to the STM. However, when assessing longitudinal traits, the MTM requires adjustment of the data for ages previously established as standards, which can cause the discarding of large numbers of records and consequent differences in the estimates of variances and genetic parameters (Nobre *et al.*, 2003b). Due to the excellent data structure in this study, no differences in the means of the parameters obtained by the MTM with standardization of weight for the ages at 330, 385, 440, 495 and 550 days were observed relative to the RRM (Figure 3).

The RRM does not require data adjustment and enables estimates of genetic parameters to be obtained at any age within the assessed range (Nobre *et al.*, 2003a; b; Sousa Júnior *et al.*, 2010). In this study, the inclusion of knots at the ages that comprised the limit of the age range, in which weight records were observed in the data file, prevented the loss of data during the analysis by this model.

The estimates of the genetic parameters for weight at each age obtained by the RRM for the DB85 and DB70 files were similar to the estimates of the genetic parameters for weight at each age for the complete file (Figure 3), which suggests the adequacy of this model for analyses of herds that practice selective data recording. Thus, the RRM becomes an alternative to the MTM since different solutions for each age of weighing and solutions for the different classes of residual variances are obtained. In addition, the use of linear polynomial splines guarantees the direct acquisition of the parameters and genetic values for each point of selection where the knot was fitted (Meyer, 2004). The use of these polynomials enables the inclusion of all weights analyzed at different ages without prior adjustment, which reduces the loss of information (Pedrosa *et al.*, 2014).

These results corroborate those obtained by Sousa Júnior *et al.* (2010), who compared the multi-trait model and random regression and obtained similar additive genetic variance estimates for both models. However, Nobre *et al.* (2003b) analyzed two data files, including complete records and incomplete data, by the multi-trait model and random regression with Legendre polynomials and observed differences in the estimates of variances and heritability among the evaluated files. The authors stated that random regression may be imprecise due to several factors, such as the size of the data set, the type of model in the analysis and the applied methodologies. The adequate size and consistency of data used in this study, the use of linear polynomial splines and the inclusion of knots at the extremities of the analyzed data enabled the RRM adequacy for the analysis of the herds that practiced selective recording.

The similarity in the genetic, residual and phenotypic correlations between the MTM and RRM for DB100, DB85 and DB70 (Tables 2, 3 and 4) suggests the suitability of both models for the computation of these estimates in the presence or absence of selective data recording. The high values obtained for this estimate suggest that the selection for weight at any age is capable of causing genetic changes in the weights at all analyzed ages. These results are consistent with the results obtained by Boligon *et al.* (2009), who evaluated the correlation among weights at birth, weaning, and two, three and five years of age for the genetic values estimated using the MTM for Nellore cattle.

Thus, we were able to verify that the use of STM can compromise the precision of the estimates of the variance components for herds that practice selective recording and that the MTM and RRM are the most appropriate models for the analysis of this type of data because they consider the correlation among the traits and enable the inclusion of the anchor trait in the analysis. The use of the RRM can simplify the genetic evaluation of longitudinal traits, as this model eliminates the need for standardizes data, directly provides the solutions of the parameters and genetic values for the fitted knots and guarantees the use of records that can be eliminated during standardization.

3.5. Conclusions

The multi-trait model and random regression model with linear spline polynomials are suitable for the genetic evaluation of longitudinal data in herds when selective data recording is employed. The inadequacy of single trait model for herds practicing sequential selection is due to selection and not the change of structure in the data file.

4.0. IMPLICATIONS

The results of this study will enhance the understanding of the behavior of genetic parameters of a longitudinal trait under selective data recording and the absence of influence of this recording in the selection process for other traits. This influence can mask what could occur with the genetic additive, residual and phenotypic variances and heritabilities for each trait separately. Thus, the multi-trait model (MTM) and random regression model with linear spline polynomials (RRM) enable accurate genetic evaluations and the identification of the animals that offer superior breeding value for a certain trait.

5.0. ACKNOWLEDGMENTS

The authors thank the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq), the Brazilian Federal Agency for Support and Evaluation of Graduate Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Capes) and the Minas Gerais Research Foundation (Fundação de Amparo à Pesquisa do Estado de Minas Gerais - Fapemig) for their financial support and the Brazilian Association of Zebu Breeders for access to the database.

6.0. REFERÊNCIAS

- ALBUQUERQUE, L. G.; MEYER, K. Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. *Journal of Animal Science*, v. 79, p. 2776–2789, 2001.
- BOHMANOVA, J.; MISZTAL, I.; BERTRAND, J. K. Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. *Journal of Animal Science*, v. 83, p. 62–67, 2005.
- BOLIGON, A. A.; ALBUQUERQUE, L. G.; MERCADANTE, M. E. Z.; LÔBO, R. B. Herdabilidades e correlações entre pesos do nascimento à idade adulta em rebanhos da raça Nelore. *Revista Brasileira de Zootecnia*, p. 2320–2326, 2009.
- BULMER, M. G. The Effect of Selection on Genetic Variability. *The American Naturalist*, v. 105, p. 201–211, 1971.
- BULMER, M. G. *The Mathematical Theory of Quantitative Genetics*. Oxford; New York: Oxford University Press, 1980.
- DIAS, L. T.; ALBUQUERQUE, L. D.; TONHATI, H.; TEIXEIRA, R. D. A. Estimação de parâmetros genéticos para peso do nascimento aos 550 dias de idade para animais da raça Tabapuã utilizando-se modelos de regressão aleatória. *Revista Brasileira de Zootecnia*, v. 35, p. 1915–1925, 2006.
- DICKERSON, G.; HAZEL, L. N. Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. *Journal Agricultural Research*, v. 69, p. 459-464, 1944.
- FELSENSTEIN, J. The Effect of Linkage on Directional Selection. *Genetics*, v. 52, p. 349–363, 1965.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian statistics*, v. 4, p. 169–193, 1992.

- GIANOLA, D.; FERNANDO, R. L. Prediction of breeding value under Henderson's selection model: a revisitation. *Journal Dairy Science*, v. 71, p. 2790–2798, 1988.
- HARRIS, D. L.; STEWART, T. S.; ARBOLEDA, C. R. Animal breeding programs: a systematic approach to their design. *Peoria : U.S.D.A.* p. 14, 1984.
- HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. *Operations Research*, v. 31, p. 1109–1144, 1983.
- HENDERSON, C. R. Estimation of changes in herd environment. *Journal of Dairy Science*, v. 32, p. 706–715, 1949.
- HENDERSON, C.R.; KEMPTHORNE, O.; SEARLE, S.R.; KROSIGK, C.M. The Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics*, v. 15, p. 192, 1959.
- HENDERSON, C. R. Sire evaluation and genetic trends. *Journal of Animal Science*, v. 1973, p. 10–41, 1973.
- HENDERSON, C. R. General Flexibility of Linear Model Techniques for Sire Evaluation. *Journal of Dairy Science*, v. 57, p. 963–972, 1974.
- HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, v. 31, p. 423, 1975.
- HENDERSON, C. R. Best Linear Unbiased Prediction of Breeding Values Not in the Model for Records. *Journal of Dairy Science*, v. 60, p. 783–787, 1977.
- HENDERSON, C. R. Recent Developments in Variance and Covariance Estimations. *Journal of Animal Science*, v. 63, p. 208–216, 1986.
- HENDERSON, C. R.; QUAAS, R. L. Multiple Trait Evaluation Using Relatives' Records. *Journal of Animal Science*, v. 43, p. 1188–1197, 1976.
- KAPS, M.; HERRING, W. O.; LAMBERSON, W. R. Genetic and environmental parameters for mature weight in Angus cattle. *Journal of Animal Science*, v. 77, p. 569–574, 1999.

- LONG, T. E.; JOHNSON, R. K.; KEELE, J. W. Effects of selection of data on estimates of breeding values by three methods for litter size, backfat, and average daily gain in swine. *Journal of Animal Science*, v. 69, p. 2787–2794, 1991.
- LUSH, J. L. Predicting gains in feeder cattle and pigs. *Journal of Agricultural Research*, v. 42, p. 853–881, 1931.
- LUSH, J. L.; SHRODE, R. R. Changes in Milk Production with Age and Milking Frequency¹. *Journal of Dairy Science*, v. 33, p. 338–357, 1950.
- MASHILOANE, M. L.; NEPHAWE, K. A.; MAIWASHE, A.; NORRIS, D. Effect of pre-weaning selection on estimates of genetic parameters and breeding values for post-weaning growth rate of beef bulls on test. *Tropical Animal Health and Production*, v. 41, p. 1723–1730, 2009.
- MERCADANTE, M. E. Z.; RAZOOK, A. G.; TROVO, J. B. F.; CYRILLO, J. N. S. G. Parâmetros genéticos do peso no início da estação de monta, considerando indicativo do peso adulto de matrizes Nelore. *Revista Brasileira de Zootecnia*, v. 33, p. 1135–1144, 2004.
- MEYER, B. K. Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. *Journal of Animal Breeding and Genetics*, v. 116, p. 181–205, 1999.
- MEYER, K. Estimates of genetic parameters for mature weight of Australian beef cows and its relationship to early growth and skeletal measures. *Livestock Production Science*, v. 44, p. 125–137, 1995.
- MEYER, K. Scope for a random regression model in genetic evaluation of beef cattle for growth. *Livestock Production Science*, v. 86, p. 69–83, 2004.
- MEYER, K. Estimates of covariance functions for growth of Angus cattle from random regression analyses fitting B-spline functions. *Proc. Assoc. Advmt. Anim. Breed. Genet.*

Anais. 2005. Disponível em: <<http://www.aaabg.org/livestocklibrary/2005/052meyer.pdf>>.

Acesso em: 13 dez. 2016

MEYER, K.; CARRICK, M. J.; DONNELLY, B. J. Genetic parameters for growth traits of Australian beef cattle from a multibreed selection experiment. *Journal of Animal Science*, v. 71, p. 2614–2622, 1993.

MISZTAL, I. Properties of random regression models using linear splines. *Journal of Animal Breeding and Genetics*, v. 123, p. 74–80, 2006.

MISZTAL, I.; TSURUTA, S.; LOURENÇO, D., AGUILAR, I., LEGARA, A.; VITEZICA, Z. *Manual for BLUPF90 family of programs*. Athens: University of Georgia. 2015.

MISZTAL, I.; LAWLOR, T. J.; SHORT, T. H. Implementation of Single- and Multiple-Trait Animal Models for Genetic Evaluation of Holstein Type Traits. *Journal of Dairy Science*, v. 76, p. 1421–1432, 1993.

NOBRE, P. R. C.; MISZTAL, I.; TSURUTA, S.; BERTRAND, J. K.; SILVA, L. O. C.; LOPES, P. S. Analyses of growth curves of Nelore cattle by multiple-trait and random regression models. *Journal of Animal Science*, v. 81, p. 918–926, 2003a.

NOBRE, P. R. C.; MISZTAL, I.; TSURUTA, S.; BERTRAND, J. K.; SILVA, L. O. C.; LOPES, P. S. Genetic evaluation of growth in Nelore cattle by multiple-trait and random regression models. *Journal of Animal Science*, v. 81, p. 927–932, 2003b.

PEDROSA, V. B.; ELER, J. P.; FERRAZ, J. B. S.; PINTO, L. F. B. Utilização de modelos unicaracterística e multicaracterística na estimação de parâmetros genéticos na raça Nelore. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v. 66, p. 1802–1812, 2014.

PHOCAS, F.; COLLEAU, J. J.; MÉNISSIER, F. Expected efficiency of selection for growth in a French beef cattle breeding scheme. I. Multistage selection of bulls used in artificial insemination. *Genetics Selection Evolution*, v. 27, p. 149–170, 1995.

- PIERAMATI, C.; VLECK, V.; D, L. Effect of genetic groups on estimates of additive genetic variance. *Journal of Animal Science*, v. 71, p. 66–70, 1993.
- POLLAK, E. J.; QUAAS, R. L. Monte Carlo Study of Genetic Evaluations Using Sequentially Selected Records. *Journal of Animal Science*, v. 52, p. 257–264, 1981.
- POLLAK, E. J.; WERF, J. VAN DER; QUAAS, R. L. Selection Bias and Multiple Trait Evaluation. *Journal of Dairy Science*, v. 67, p. 1590–1595, 1984.
- RAFTERY, A. E.; LEWIS, S. M. One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Statistical Science*, v. 7, p. 492–497, 1992.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org/>.
- SAKAGUTI, E. S.; SILVA, M. A.; QUAAS, R. L.; MARTINS, E. N.; LOPES, P. S.; SILVA, L. O. C. Avaliação do crescimento de bovinos jovens da raça Tabapuã, por meio de análises de funções de covariâncias. *Revista Brasileira de Zootecnia*, v. 32, p. 864–874, 2003.
- SARMENTO, J. L. R.; TORRES, R. A.; SOUSA, W. H.; PEREIRA, C. S.; LOPES, P. S.; BREDA, F.C. Estimação de parâmetros genéticos para características de crescimento de ovinos Santa Inês utilizando modelos uni e multicaracterísticas. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v. 58, p. 581–589, 2006.
- SCHAEFFER, L. R. Sire and Cow Evaluation Under Multiple Trait Models. *Journal of Dairy Science*, v. 67, p. 1567–1580, 1984.
- SCHAEFFER, L. R.; SCHENKELLY, F. S.; FRIES, L. A. Selection bias on animal model evaluation. *Journal of Animal Science*, v. 41, 1997.
- SMITH, B. J. Bayesian output analysis program (BOA) version 1.1 user's manual. Dept. of Biostatistics, Univ. of Iowa, College of Public Health. Disponível em: [http://www. public-health. uiowa. edu/boa](http://www.public-health.uiowa.edu/boa), 2005.

SORENSEN, D. A.; KENNEDY, B. W. Estimation of Response to Selection Using Least-Squares and Mixed Model Methodology. *Journal of Animal Science*, v. 58, p. 1097–1106, 1984a.

SORENSEN, D. A.; KENNEDY, B. W. Estimation of Genetic Variances from Unselected and Selected Populations. *Journal of Animal Science*, v. 59, p. 1213–1223, 1984b.

SORENSEN, D. A.; KENNEDY, B. W. Analysis of Selection Experiments Using Mixed Model Methodology. *Journal of Animal Science*, v. 63, p. 245–258, 1986.

SORENSEN, D.; GIANOLA, D. Likelihood, bayesian, and MCMC methods in quantitative genetics. *Springer*, 2002.

SOUSA, W. H.; PEREIRA, C. S.; BERGMANN, J. A. G.; SILVA, F. L. R. Estimativas de componentes de (co) variância e herdabilidade direta e materna de pesos corporais em ovinos da raça Santa Inês. *Revista Brasileira de Zootecnia*, v. 28, p. 1252–1262, 1999.

SOUSA JÚNIOR, S. C. S.; OLIVEIRA, S. M. P; ALBUQUERQUE, L. G; BOLIGON, A. A; FILHO, R. M. Estimação de funções de covariância para características de crescimento da raça Tabapuã utilizando modelos de regressão aleatória. *Revista Brasileira de Zootecnia*, p. 1037–1045, 2010.

TASSELL, C. P. V.; CASELLA, G.; POLLAK, E. J. Effects of Selection on Estimates of Variance Components Using Gibbs Sampling and Restricted Maximum Likelihood. *Journal of Dairy Science*, v. 78, p. 678–692, 1995.

THOMPSON, R.; MEYER, K. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science*, v. 15, p. 299–313, 1986.

UNITED STATES DEPARTMENT OF AGRICULTURE (USDA). Disponível em:

<http://www.usda.gov/wps/portal/usda/usdahome>> Acesso em: 25 nov. 2016

VALVERDE, R. R.; ALVAREZ, O. C. H.; DOMINGUEZ, R. N.; FLORES, A. R.; MUÑIZ, J. G. G. Análisis univariado vs multivariado en la evaluación genética de variables de crecimiento en dos razas bovinas. *Agrociencia*, v. 41, p. 271–282, 2007.

WALTER, J. P.; MAO, I. L. Multiple and Single Trait Analyses for Estimating Genetic Parameters in Simulated Populations Under Selection. *Journal of Dairy Science*, v. 68, p. 91–98, 1985.

YATES, F. The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes. *Journal of the American Statistical Association*, v. 29, p. 51, 1934.