

**UNIVERSIDADE FEDERAL DO PAMPA**

**CHARLES RODRIGUES BASTOS**

**APLICAÇÃO DE MODELOS ROBUSTOS  
PARA A PREDIÇÃO DE VALORES  
GENÉTICOS EM BOVINOS DE CORTE**

**Bagé  
2020**

**CHARLES RODRIGUES BASTOS**

**APLICAÇÃO DE MODELOS ROBUSTOS  
PARA A PREDIÇÃO DE VALORES  
GENÉTICOS EM BOVINOS DE CORTE**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Orientador: Fernando Flores Cardoso  
Co-orientador: Sandro da Silva Camargo

**Bagé  
2020**

Ficha catalográfica elaborada automaticamente com os dados fornecidos pelo autor através do Módulo de Biblioteca do Sistema GURI (Gestão Unificada de Recursos Institucionais).

B327a Bastos, Charles Rodrigues

Aplicação de modelos robustos para a predição de valores genéticos em bovinos de corte / Charles Rodrigues Bastos.

79 p.

Dissertação (Mestrado) - Universidade Federal do Pampa, PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA, 2020.

"Orientação: Fernando Flores Cardoso;  
Co-orientação: Sandro da Silva Camargo".

1. Predição genética. 2. Valores extremos.  
3. Melhoramento genético animal. 4. Acurácia.  
5. Modelo robusto. I. Título.

**CHARLES RODRIGUES BASTOS**

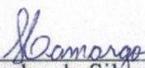
**APLICAÇÃO DE MODELOS ROBUSTOS  
PARA A PREDIÇÃO DE VALORES  
GENÉTICOS EM BOVINOS DE CORTE**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Área de concentração: Tecnologias para a produção agropecuária

Dissertação defendida e aprovada em: 03 de setembro de 2019.

Banca examinadora:



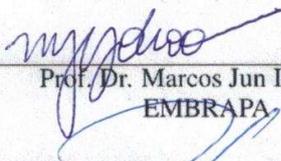
---

Prof. Dr. Sandro da Silva Camargo  
Co-orientador



---

Prof. Dr. Arione Augusti Boligon  
UFPEL



---

Prof. Dr. Marcos Jun Iti Yokoo  
EMBRAPA



---

Prof. Dr. Paulo Fernando Marques Duarte Filho  
UNIPAMPA

Dedico este trabalho aos meus filhos  
Amanda e Raul.

## **AGRADECIMENTO**

À UNIPAMPA, por proporcionar a continuidade da minha formação acadêmica oferecendo um programa de pós-graduação de alto nível, público e gratuito.

Aos meus orientadores, Fernando e Sandro, pelo ensinamento, orientações e apoio durante esta jornada.

Ao professor Marcos Yokoo, por todo o apoio e atenção durante este trabalho.

Aos professores do PPGCAP, pelo ensinamento e experiências acadêmicas compartilhadas.

Aos colegas do PPGCAP, pelo companheirismo e por todos os bons momentos vivenciados.

À minha esposa, pela compreensão e pela disposição.

Aos meus pais, pelo estímulo e encorajamento.

Aos meus irmãos, que também estão em busca de seus objetivos acadêmicos, pela torcida e pelo incentivo mútuo.

Aos meus filhos, pelo precioso tempo que 'emprestamos' a este projeto.

“O preferível não é o desejo de acreditar, mas o desejo de descobrir, que é exatamente o oposto.”

— Bertrand Russell

## RESUMO

Valores extremos podem distorcer o resultado de uma avaliação genética. Da mesma forma, a exclusão destes valores pode ocultar alterações relevantes em um rebanho. A predição dos valores genéticos, em uma população de indivíduos, deverá ter um nível mais elevado de precisão quando a informação fenotípica e de pedigree disponíveis corresponderem a dados fidedignos. Entretanto, fatores como o efeito potencial de lesões desconhecidas, doenças, tratamento diferenciado ou até mesmo erros de entrada de dados são variáveis que não são consideradas nos modelos estatísticos, mas são capazes de comprometer a qualidade dos dados a ponto de influenciar significativamente o desempenho de um indivíduo, ou grupo de indivíduos, gerando valores extremos que poderão enviesar as estimativas dos parâmetros genéticos. Os modelos estatísticos mistos são os mais utilizados para a predição de valores genéticos porém, são sensíveis a dados com valores extremos e necessitam editar ou descartar estes dados, para mitigar a distorção dos resultados. Diante disso, o objetivo deste trabalho é demonstrar que a implementação de um modelo robusto pode reduzir a influência destes dados com valores extremos e melhorar o resultado da predição sem descartar dados. Para tanto, foi desenvolvido um algoritmo que calcula as equações de modelos mistos, identifica a relação entre os valores extremos e o resultado da predição, introduzindo, quando necessário, uma variável de ponderação capaz de reduzir o desvio de cada observação em relação à média de sua unidade amostral. Os resultados obtidos demonstraram que foi possível melhorar a precisão das estimações, reduzindo, em alguns casos, a influência de valores extremos em até 90 por cento, de acordo com o desvio padrão calculado, sem descartá-los do modelo. Desta forma, diante de conjuntos de dados com valores extremos, o modelo robusto de predição apresentou resultados mais precisos, em comparação ao modelo misto. Nas duas características avaliadas, houve reduções entre 55 e 79 por cento no intervalo entre o maior e o menor valor estimado.

**Palavras-chave:** Predição genética. Valores extremos. Melhoramento genético animal. Acurácia. Modelo robusto.

## ABSTRACT

Extreme values can distort the result of a genetic assessment. Similarly, deleting these values can hide relevant changes in a herd. Prediction of genetic values in a population of individuals should have a higher level of accuracy when the available phenotypic and pedigree information matches reliable data. However, factors such as the potential effect of unknown injuries, disease, differential treatment or even data entry errors are variables that are not considered in statistical models, but are capable of compromising data quality to the extent that they significantly influence performance of an individual or group of individuals, generating extreme values that may skew estimates of genetic parameters. Mixed statistical models are the most used for predicting genetic values, but are sensitive to data with extreme values and need to edit or discard these data to mitigate the distortion of results. Therefore, the objective of this work is to demonstrate that the implementation of a robust model can reduce the influence of this data with extreme values and improve the prediction result without discarding data. For this, an algorithm was developed that calculates the mixed model equations, identifies the relationship between the extreme values and the accuracy of the prediction and introduces, when necessary, a weighting variable capable of reducing the deviation of each observation from the mean. The results showed that it was possible to improve the accuracy of the estimates, reducing, in some cases, the influence of extreme values by up to 90 percent, according to the calculated standard deviation, without discarding them from the model. Thus, in the face of data sets with extreme values, the robust prediction model presented more accurate results compared to the mixed model. In both characteristics evaluated, there were reductions between 55 and 79 percent in the interval between the highest and the lowest estimated value.

**Keywords:** Genetic prediction. Extreme values. Animal genetic improvement. Accuracy. Robust model.

## LISTA DE FIGURAS

Figura 1	Resultados do BLUP para ganho de peso à desmama em bovinos de corte.....	26
Figura 2	Predição de valores genéticos em conjuntos com <i>outliers</i> .....	35
Figura 3	Oscilação das médias dos valores genéticos, para machos e fêmeas, a cada iteração.....	38
Figura 4	Oscilação dos valores genéticos individuais, a cada iteração. ....	38
Figura 5	Histograma dos valores fenotípicos para as características AOL e EG.....	47
Figura 6	Fluxograma das etapas do processo de cálculo dos valores genéticos e mitigação dos efeitos causados por dados extremos.....	48
Figura 7	Dispersão dos valores genéticos e desvios padrão após o cálculo da MME. ....	51
Figura 8	Dispersão dos valores genéticos recalculados com a matriz de ponderação. ....	52
Figura 9	<i>Outliers</i> se aproximando da linha principal.....	53
Figura 10	Dispersão dos valores genéticos de todo o conjunto de indivíduos após o processamento.....	54
Figura 11	Predição dos valores genéticos do animal id = 1183 a cada iteração. ....	54
Figura 12	Predição das médias dos valores genéticos separadas por grupos contemporâneos. ....	55
Figura 13	Predição das médias dos valores genéticos do grupo contemporâneo 03. ....	56
Figura 14	Predição das médias dos valores genéticos do grupo contemporâneo 12. ....	56
Figura 15	Predição das médias dos valores genéticos do grupo contemporâneo 29. ....	56
Figura 16	Maior e menor valores individuais para a característica EG. ....	57
Figura 17	Dispersão dos valores genéticos de EG após cada iteração.....	58
Figura 18	Histogramas dos valores individuais preditos para EG. ....	59
Figura 19	Média do valor genético da AOL estimado para cada grupo contemporâneo.....	61
Figura 20	Correlação entre os valores genéticos individuais obtidos e a prova de conceito. ....	61
Figura 21	Comparativo do tempo de processamento para o cálculo das MME.....	64

## LISTA DE TABELAS

Tabela 1	Ganho de peso, à desmama, para bezerros de corte. ....	25
Tabela 2	Pedigree para bezerros de corte .....	26
Tabela 3	Matriz de consanguinidade entre os indivíduos .....	27
Tabela 4	Procedimento para inversão da matriz (A) .....	31
Tabela 5	Matriz inversa de consanguinidade entre os indivíduos ( $A^{-1}$ ) .....	31
Tabela 6	Dados de WWG com <i>outlier</i> .....	33
Tabela 7	Predição de valores genéticos com o BLUP, considerando os valores extremos	34
Tabela 8	Matriz de ponderação ( <i>D</i> ) em cada iteração .....	37
Tabela 9	Análise descritiva dos dados .....	46
Tabela 10	Valores genéticos de <i>outliers</i> estimados sem matriz e com matriz de ponderação. ....	52
Tabela 11	Comparação das médias dos valores estimados para EG após a última iteração. ....	59
Tabela 12	Resumo dos valores da acurácia das predições. ....	62
Tabela 13	Correlação entre a acurácia da predição calculada, a acurácia da primeira e última iterações e a acurácia calculada no BRBV. ....	63
Tabela 14	Correlação entre a predição inicial do modelo, sua primeira e última iterações e a prova de conceito. ....	63

## LISTA DE ABREVIATURAS E SIGLAS

ABIEC	Associação Brasileira das Indústrias Exportadoras de Carnes
AOL	Área de Olho de Lombo
BLUP	<i>Best Linear Unbiased Prediction</i>
BRBV	<i>Brazilian Breeding Values</i>
CTBS	<i>Constrained Translated Biweight S-estimator</i>
DEP	Diferença Esperada na Progenie
DP	Desvio Padrão
EG	Espessura de Gordura
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
GC	Grupo Contemporâneo
GBLUP	<i>Genomic Best Linear Unbiased Prediction</i>
MAP	<i>Maximum A posteriori Prediction</i>
MCMC	<i>Markov chain Monte Carlo</i>
MLE	<i>Maximum likelihood estimator</i>
MME	<i>Mixed Model Equations</i>
PIB	Produto Interno Bruto
REML	<i>Restricted Maximum Likelihood</i>
WWG	<i>Weaning Weight Gain</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
1.1 Justificativa .....	16
1.2 Problema de Pesquisa .....	17
1.3 Objetivos .....	17
1.4 Organização do trabalho .....	18
<b>2 CONCEITOS GERAIS E REVISÃO DE LITERATURA</b> .....	<b>19</b>
2.1 Seleção Genética.....	19
2.2 Modelos genéticos estatísticos .....	21
2.3 Inferência com o BLUP .....	22
2.4 Matriz de Parentesco .....	26
2.5 Modelos de predição robusta .....	31
2.6 Matriz de Ponderacao.....	34
2.7 Trabalhos Correlatos .....	39
<b>3 METODOLOGIA</b> .....	<b>44</b>
3.1 Caracterização da pesquisa.....	44
3.2 Material e métodos.....	45
<b>4 RESULTADOS E DISCUSSÃO</b> .....	<b>50</b>
4.1 Predição dos valores genéticos da área de olho de lombo (AOL) .....	50
4.2 Predição dos valores genéticos da espessura de gordura (EG).....	57
4.3 Validação dos resultados .....	59
4.4 Acurácia .....	61
4.5 Avaliação de desempenho .....	63
<b>5 CONCLUSÕES</b> .....	<b>66</b>
<b>REFERÊNCIAS</b> .....	<b>67</b>
<b>APÊNDICE A — CÁLCULO DO BLUP PARA A PREDIÇÃO DE GANHO DE PESO À DESMAMA E EFEITOS FIXOS DO SEXO</b> .....	<b>72</b>
<b>APÊNDICE B — VALORES GENÉTICOS ESTIMADOS PARA A CARAC- TERÍSTICA DE AOL, SUBMETIDOS À PROVA DE CONCEITO</b> .....	<b>78</b>

## 1 INTRODUÇÃO

A pecuária brasileira se estabeleceu, nos últimos anos, como uma das mais relevantes do atual cenário mundial, devido ao volume de exportações (ROSA, 2018). É fato, também, que a pecuária contribui de maneira expressiva para o crescimento da economia nacional, sendo fundamental para a manutenção do saldo comercial positivo nos últimos anos (ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNES, 2018). O índice macroeconômico utilizado para mensurar esta contribuição é o Produto Interno Bruto (PIB), que é utilizado para representar a soma de todos os bens e serviços produzidos em uma determinada região, durante um determinado período (CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL, 2018). Do ponto de vista financeiro, de acordo com Associação Brasileira das Indústrias Exportadoras de Carnes (ABIEC) o agronegócio corresponde, atualmente, a cerca de 22% do PIB brasileiro e a pecuária é responsável por 31% deste montante. No ano de 2017, as exportações brasileiras de carne bovina encerraram o ano com uma receita de US\$ 6,2 bilhões, um crescimento de 13% em relação ao ano anterior (ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNES, 2018). Em 2018, o Brasil quebrou o recorde de exportação, acumulando um novo crescimento financeiro de aproximadamente 8% frente ao resultado de 2017 (ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNES, 2019). Já pelo ponto de vista do volume, as exportações brasileiras de carne bovina atingiram 1,64 milhão de toneladas em 2018, volume 11% superior ao registrado em 2017. Diante disso, a Confederação da Agricultura e Pecuária do Brasil (CNA) apontou o setor agrícola como responsável por 44% do total das exportações brasileiras, além de ser responsável pela geração de 32% de todos os empregos existentes, atualmente, no país (CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL, 2018).

O crescimento das exportações de carne bovina nos últimos anos, destacado pela ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNES (2018), reflete a capacidade brasileira de acompanhar a demanda por produtos cada vez mais qualificados, capazes de atender os rigorosos níveis de exigência do mercado internacional. Atualmente são exigidos diversos requisitos sanitários, controle alimentar e de vacinações além de aspectos quanto à qualidade da carne, destacados por Bridi (2004), que podem envolver características como: cor, quantidade e espessura de gordura, valor nutricional e ausência de resíduos. Polaquini, Souza e Gebara (2006) destacam que

contribuíram para intensificar as melhorias fatores como: ajustes internos de qualidade e aspectos sanitários, entre outros, que são relevantes para a expansão das exportações de carne bovina. Diante disso, a preocupação em elevar a posição brasileira no cenário mundial fez com que a adoção de programas de melhoramento animal se tornasse um diferencial competitivo.

Apesar dos avanços comerciais dos últimos anos, Vicensotti, Montebello e Marjotta-Maistro (2019) consideram que a concorrência do mercado internacional faz com que exista a necessidade de monitoramento e aprimoramento constante de todos os processos que envolvem a cadeia produtiva da pecuária brasileira. Para tanto, são analisados os índices zootécnicos e econômicos, definidos por Euclides Filho (2000) como indicadores da eficiência dos sistemas de produção, capazes de sinalizar erros e sugerir alterações pontuais possibilitando o aprimoramento dos processos. Neste sentido, de acordo com Vicensotti, Montebello e Marjotta-Maistro (2019), o destaque brasileiro no cenário mundial é decorrência de fatores positivos como: a ampla extensão territorial, as fontes alimentícias apropriadas para a criação de gado, o melhoramento genético para adaptação dos animais ao meio ambiente e o emprego de tecnologias capazes de aumentar os índices de produtividade.

A necessidade de satisfazer os padrões de exigência do mercado consumidor e ainda propiciar lucro ao produtor, é um desafio que motiva a busca por tecnologias capazes de proporcionar a adaptação de rebanhos em diferentes ambientes ou potencializar determinadas características, comercialmente importantes, em um rebanho (ROSA, 2018). Diante deste contexto, o aprimoramento genético é uma das preocupações dos programas de melhoramento animal e pode ser otimizado através da avaliação acurada dos valores genéticos e fenotípicos (EUCLIDES FILHO, 2000). Outras medidas como: modificações no manejo, tratamento e alimentação também podem gerar resultados eficazes. Entretanto, somente o melhoramento genético caracteriza-se pela promoção de alterações hereditárias, ou seja, garante que os aprimoramentos serão transmitidos para as próximas gerações, tornando os ganhos permanentes e acumulativos (STRANDÉN; GIANOLA, 1999).

O melhoramento genético consiste em alterar a proporção de certos genes, com o objetivo de maximizar a expressão de alguma característica desejada no animal, normalmente com o intuito de obter maior lucratividade para o dono do rebanho. Isso é feito através da seleção e direcionamento dos acasalamentos. Para proceder com a seleção e obter maiores ganhos, um parâmetro importante é a herdabilidade da característica em

questão. considerando os índices de herdabilidade. Bourdon e Bourbon (2000) define a herdabilidade como um coeficiente de herança, de uma dada característica, que pode ser classificada de acordo com sua magnitude como: baixa, moderada ou alta, conforme a correlação entre o genótipo, que é patrimônio genético do animal, e fenótipo, que são características observáveis resultantes da interação do genótipo com o ambiente. Desta forma, a baixa herdabilidade sugere que o desempenho apresentado por um animal está associado ao ambiente a ele proporcionado. Entretanto, herdabilidades mais altas sugerem que são maiores as probabilidades do desempenho de um animal refletir sua genética. Em geral, quanto maior a herdabilidade de uma característica, maior será a acurácia da seleção e maior a resposta à seleção (PEREIRA, 2004).

O ganho genético, de acordo com Pereira (2004), é o parâmetro que exprime o avanço da geração seguinte em relação ao rebanho original, decorrente da seleção efetuada. A utilização de ferramentas para estudo populacional, mapeamento e análises de similaridade, distância genética e predição de valores fenotípicos, pode viabilizar um aumento nas taxas anuais de ganho genético. Em alguns países, essas taxas variam entre 1 e 2% (MONTALDO; MEZA-HERRERA, 1998). Este incremento, aos poucos, está sendo introduzido na pecuária nacional através dos programas de melhoramento animal que, conforme Euclides Filho (2000), são serviços especializados, capazes de subsidiar o criador na utilização dos recursos genéticos para a produção de genética superior. Programas como: o GenePlus (ROSA, 2018) e o PampaPlus (CARDOSO; LOPA; TEIXEIRA, 2016) vêm proporcionando resultados significativos e, por consequência, despertando o interesse de produtores e associações de raças.

Para estimar o ganho genético, Rosa (2018) apresenta uma série de regras para a coleta e análise de consistência dos dados. A avaliação considera o uso de um modelo genético estatístico, detalhado na seção 2.2, enquanto os resultados são valores genéticos, obtidos para cada animal, expressos em Diferença esperada na progênie (DEP), detalhada na seção 2.1. Os modelos genéticos estatísticos possibilitam ao pesquisador prever resultados capazes de indicar se haverá, ou não, evolução de uma ou mais características em um conjunto de indivíduos monitorados (PEREIRA, 2004). Neste contexto, os modelos mistos definidos por Henderson (1973) são implementados em diversas metodologias, com destaque para o *Best Linear Unbiased Prediction* (BLUP) (HENDERSON, 1977) que é o modelo misto mais difundido atualmente. Por definição, os modelos mistos abordam tanto os efeitos fixos, ou imutáveis, tais como o sexo ou data do parto, quanto os efeitos aleatórios, ou incertos, tais como peso e altura (GIANOLA; IM; MACEDO,

1990).

A resolução do BLUP é obtida mediante um sistema de equações lineares, definido como *Mixed Model Equations* (MME) que, na prática, é viabilizado pela utilização de soluções computacionais (MRODE, 2014). De acordo com Jiang (2007), a resolução do BLUP é dificultada pelas dimensões do sistema de equações que precisa ser resolvido e pela sensibilidade a valores extremos que, quando computados, tendem a enviesar os resultados da análise. De acordo com Mrode (2014), a resolução de equações com grandes volumes de dados pode gerar centenas de milhares de cálculos e, conforme Barbosa, Pereira e Oliveira (2018), a identificação visual dos valores extremos é impraticável nestes casos, fato que justifica a exclusão de valores, muito afastados da média geral, para que não exerçam influência no resultado da análise. De maneira geral, são considerados extremos, os valores fenotípicos afastados da normalidade em mais de 4 desvios padrão da média.

A mitigação das interferências causadas pelos valores extremos, ou *outliers*, é abordada por Cardoso *et al.* (2009) através da utilização de modelos que não são sensíveis a *outliers*, denominados modelos robustos. Estes modelos se caracterizam pela capacidade de analisar variações, dentro de uma unidade amostral, considerando os dados com valores extremos, que seriam desprezados nas equações de modelos mistos pelo fato de estarem muito afastados da média. Para moderar as distorções no resultado final, o modelo robusto considera a inclusão de uma matriz de ponderação que será preenchida com valores proporcionais à média da unidade amostral, obtidos de equação específica, cuja finalidade consiste em redefinir o fator de influência que um indivíduo com valor extremo será capaz de exercer, na avaliação da média do seu grupo contemporâneo.

## 1.1 Justificativa

Para a definição de indivíduos geneticamente superiores é necessária a reunião de uma série de atributos favoráveis, que lhes confirmam rendimento comparativamente mais elevado em relação aos demais indivíduos do rebanho (CRUZ, 2008). A avaliação das características genéticas permite a identificação de animais geneticamente superiores, que poderão transmitir essa superioridade para as próximas gerações. Atualmente, o método mais utilizado pelos pesquisadores para predição de valores genéticos é o modelo de equações mistas denominado BLUP que, comparado a outros modelos mistos, apresenta os resultados mais acurados. Entretanto, o BLUP é sensível à presença de *outliers*

e sofre distorções nos resultados das análises se não houver intervenção do pesquisador para tratar, ou simplesmente excluir, os dados com valores mais extremos (YAMANAKA *et al.*, 2018). Diante disso, pretende-se demonstrar que a inferência dos valores genéticos através de um modelo robusto é capaz de proporcionar resultados mais acurados, em relação à inferência com o BLUP, em um conjunto de dados com *outliers*. Justifica-se o presente estudo, pelo fato de que a eficiência de um programa de melhoramento genético depende da realização de avaliações com resultados fidedignos e, normalmente, estes resultados são obtidos através do processamento e análise de conjuntos massivos de dados, o que inviabiliza a verificação individual e possibilita que dados fenotípicos com valores desproporcionais, à média geral do rebanho, consigam enviesar o resultado da análise (PEREIRA *et al.*, 2014). Esta abordagem foi aplicada sobre dados reais, de animais pertencentes ao programa de melhoramento animal da EMBRAPA Pecuária Sul, e os resultados obtidos servirão como contribuição para a otimização das pesquisas realizadas pela EMBRAPA. Da mesma forma, o algoritmo desenvolvido poderá colaborar com o aprimoramento dos processos de melhoramento genético na EMBRAPA.

## 1.2 Problema de Pesquisa

É possível melhorar a acurácia das estimações de valores genéticos, realizadas através do cálculo das equações de modelos mistos, em conjuntos de dados que possuem valores extremos, ou dispersos da média geral, sem desprezar ou alterar registros válidos?

## 1.3 Objetivos

O objetivo deste trabalho é demonstrar que a aplicação de um modelo robusto é capaz de reduzir a influência de valores extremos, em uma predição de valores genéticos para uma população de bovinos de corte.

De acordo com o objetivo geral estabelecido, o trabalho é planejado considerando os seguintes objetivos específicos:

- Identificar a relação entre os valores extremos e a precisão da estimação dos efeitos fixos;
- Implementar os modelos mistos, para avaliar as suas principais características e comparar os resultados;

- Demonstrar que a introdução da matriz de ponderação é capaz de reduzir o desvio de cada observação, em relação ao verificado nos modelos mistos.

#### **1.4 Organização do trabalho**

Este trabalho está organizado da seguinte forma: O capítulo 2 trata dos conceitos gerais e revisão de literatura, onde são abordados os modelos de inferência em genética quantitativa, apresentando os principais aspectos dos modelos genéticos, as características gerais dos modelos mistos e a implementação do BLUP. Também são abordadas as técnicas de ponderação para reajustar a influência de valores extremos, adotadas pelos modelos robustos, como alternativa para não desprezar dados válidos. E, ao final, os trabalhos correlatos. No capítulo 3, é mostrada a metodologia utilizada para a realização deste trabalho, a descrição dos procedimentos metodológicos que classificam a presente pesquisa, o material e os métodos utilizados para a obtenção dos resultados. No capítulo 4, são apresentados os resultados das análises realizadas e as diferenças entre as estimações realizadas com os diferentes métodos. Também são abordadas as validações dos resultados obtidos, o cálculo da acurácia e o desempenho computacional. No capítulo 5, são apresentadas as conclusões.

## 2 CONCEITOS GERAIS E REVISÃO DE LITERATURA

Neste capítulo serão apresentados os conceitos gerais acerca dos elementos que conduzem o melhoramento animal. Inicialmente serão abordados os principais aspectos da seleção genética e os modelos genéticos estatísticos mais relevantes, com destaque para o modelo animal. Em seguida, o método tradicional de inferência genética será apresentado através do BLUP, que é o modelo linear mais utilizado para predição de valores genotípicos. Também será apresentada a matriz de parentesco, que é usada pelo BLUP para identificar relações de consanguinidade. É mencionado, ainda, o sistema *Brazilian Breeding Values* (BRBV), que foi utilizado como prova de conceito para validar os resultados obtidos com o BLUP. Ao final do capítulo será apresentada a estratégia para predição robusta, que tem como característica principal a manutenção de valores extremos que normalmente seriam desconsiderados no método tradicional e, a matriz de ponderação que possibilita reduzir a influência dos valores extremos mantidos pelo modelo robusto.

### 2.1 Seleção Genética

O melhoramento genético de animais é um processo que tem por objetivo ampliar a proporção de uma determinada característica, tal como ganho de peso diário, dentro de um rebanho, geralmente visando atender o interesse econômico. Para tanto, a seleção e o sistema de acasalamento são as etapas, tradicionalmente utilizadas, para promover alterações na constituição genética de uma população (LÔBO; LÔBO, 2007). Pollack (1992) define as duas etapas da seguinte forma: a primeira consiste em identificar indivíduos com determinados valores genéticos mais elevados; e a segunda, que é a etapa da reprodução, onde reprodutores, ou touros, e vacas, ou matrizes, identificados como geneticamente superiores, são indicados para o acasalamento, ou reprodução. Esta indicação, na prática, sugere que, dentre os indivíduos avaliados, dadas as possíveis restrições, a maior probabilidade de ampliar a característica desejada, na próxima geração, é através deste par. Desta forma, conforme Lôbo e Lôbo (2007), é priorizada a reprodução de indivíduos com características superiores, permitindo que animais com um genótipo mais qualificado possam produzir uma prole maior. Este processo pode ser repetido durante algumas gerações até que os genes desejados se tornem frequentes dentro de uma população. O Valor genético, de acordo com Pollack (1992), é a parte dos componentes genéticos cuja transmissão se consegue prever e, dessa forma, avaliar o impacto na composição genética das gerações

futuras. De maneira geral, como mostra a Fórmula 1, as características expressam valores fenotípicos ( $P$ ) que são resultantes do seu valor genotípico ( $G$ ) somados aos desvios de ambiente ( $E$ ).

$$P = G + E \quad (1)$$

Onde:  $P$  é o valor fenotípico, correspondente à soma de  $G$ , que é a característica genética herdada dos pais e,  $E$ , que é o efeito que os fatores externos, ou ambientais, exercem sobre a característica do animal. Considerando que cada um dos pais transmite, aleatoriamente, por meio dos gametas, metade de seus cromossomos aos filhos, a DEP, segundo Rosa (2018), é a predição da média do valor genético de todos os gametas produzidos por um animal, quando acasalado com animais que possuam um potencial genético equivalente, comparado com um grupo de animais referências que tenham participado da mesma avaliação genética. Na prática, a DEP prediz a capacidade de transmissão genética de um animal. O conhecimento da DEP permite a tomada de decisões capazes de proporcionar contínuas evoluções genéticas e aumento da produtividade. Normalmente a DEP é expressa na unidade de medida original da característica em questão e serve para comparar o valor genético de animais, para várias características, e predizer a habilidade de transmissão genética de um animal avaliado como progenitor. Desta forma, em uma avaliação é esperada uma mesma proporção de animais com valores positivos e negativos para cada característica. De acordo com Pereira (2004), as etapas para a realização da avaliação genética são as seguintes:

- Coleta de dados: Procedimento de apuração dos valores fenotípicos em épocas preestabelecidas, tais como: peso ao desmame (205 dias), peso ao ano (365 dias), peso ao sobreano (550 dias);
- Envio dos dados: procedimento de submeter os dados coletados ao centro de avaliação, local onde os dados serão armazenados em um banco de dados e serão utilizados para pesquisas;
- Formação de grupos contemporâneos: consiste em agrupar indivíduos de características semelhantes, tais como: ano, sexo, raça, estação de nascimento e lote de manejo;
- Montagem dos arquivos de dados que serão utilizados nas análises;
- Cálculo do grau de parentesco de todos os indivíduos;
- Definição do modelo genético estatístico e formação das equações para solução simultânea; e

- Produção dos resultados da avaliação genética.

## 2.2 Modelos genéticos estatísticos

Modelos genéticos estatísticos são procedimentos que fornecem os parâmetros relevantes para a identificação de genótipos superiores (MAIA *et al.*, 2011). Dentre os métodos conhecidos para predição de valores genéticos, comparados pelo estudo de Faria *et al.* (2007), os modelos genéticos estatísticos são os que apresentam soluções mais acuradas. Diversos modelos conhecidos são utilizados para as avaliações genéticas, os mais comuns são: o modelo touro, o modelo touro-avó-materno, o modelo touro-vaca, o modelo animal e o modelo animal reduzido. A decisão de qual modelo será utilizado depende dos objetivos do pesquisador, do tipo e número de informações e dos recursos computacionais disponíveis. De acordo com Ferraz e Eler (2010), os modelos mais relevantes são: o modelo touro (*sire model*), onde apenas os touros têm uma equação, e o modelo animal (*animal model*), que considera uma equação para cada indivíduo. O modelo fixo, que antecedeu a ambos e foi utilizado até o final dos anos 80, empregava um método denominado mérito genético, que consistia na diferença média da progênie de um indivíduo para a média da sua raça, como indicação do valor genético de um touro. Sem considerar, para efeito de cálculos, os efeitos aleatórios.

Diferente do modelo fixo, os modelos touro e animal são classificados como mistos, pois consideram, simultaneamente, efeitos fixos e aleatórios. De acordo com Pereira (2004), a diferença entre eles se dá na maneira como são propostas as equações e, por consequência, a produção dos resultados da avaliação genética, sendo:

- Modelo touro: Considera a inclusão dos dados das progênies, no sistema de equações, para cada um dos respectivos pais. Como resultado, fornece a DEP apenas para touros;
- Modelo animal: Utiliza informações disponíveis de reprodutores, matrizes e produtos, assim como a relação de consanguinidade entre eles. O resultado obtido é mais preciso e são conhecidos os valores genéticos e reprodutivos de cada indivíduo.

Desta forma, de acordo com Amaral, Sereno e Pellegrin (2009), o modelo animal apresenta duas grandes vantagens sobre o modelo touro. A primeira é o fornecimento de DEP para todos os indivíduos, inclusive aos que não têm dados próprios, através de seus ascendentes, descendentes ou colaterais. A outra vantagem é a possibilidade de comparar,

para indicação de acasalamentos, animais nascidos em épocas distintas. Isso aumenta o grupo de indivíduos possíveis de serem selecionados e tornou-se possível devido a inclusão do valor genético das vacas no modelo. Apesar de ser a modelagem cujos resultados apresentavam maior acurácia, ou seja, eram mais aproximados dos valores reais, a demanda por recursos computacionais foi um limitador para a utilização do modelo animal até o final dos anos 80 (FERRAZ; ELER, 2010).

A acurácia, conforme Cardoso, Lopa e Teixeira (2016), é a correlação entre o valor genético real e o valor genético predito a partir das informações genótípicas ou fenotípicas dos indivíduos e reflete o grau de confiança depositada na precisão da estimativa da DEP. A predição acurada do valor genético dos indivíduos depende, em grande parte, dos efeitos considerados no modelo estatístico utilizado para a avaliação dos animais. Aspectos alimentares, climáticos, de manejo, período de nascimento, entre outros, são associados ao ambiente e podem ser minimizados por meio da inclusão dos grupos contemporâneos, que são definidos por Bourdon e Bourbon (2000), como um agrupamento de indivíduos que são manejados em condições ambientais semelhantes, visando reduzir os efeitos do meio sobre a expressão das características de interesse. Yokoo (2005) classifica os grupos contemporâneos por: ano, sexo, estação de nascimento e lote de manejo.

Os fatores genéticos como: grupamento, raça e parâmetros intrínsecos, como idade adequada para coleta de dados e curvas de crescimento, depois de estabelecidos, devem ser mensurados para que possam contribuir com a acurácia da predição dos valores genéticos. Um estimador acurado, de acordo com Resende *et al.* (2012), é aquele que apresenta a menor diferença quadrática entre valores reais e os valores estimados, ou seja, apresenta mínimo erro quadrático médio (EQM) de estimação. O estudo de Resende *et al.* (1996) comparou vários métodos para predição de valores genéticos, considerando o resultado do BLUP equivalente ou superior em todos os casos, quando comparado com os seguintes métodos: *Ordinary Least Squares* (OLS), *Generalized Least Squares* (GLS), *Best Prediction* (BP) e *Best Linear Prediction* (BLP).

### **2.3 Inferência com o BLUP**

O método de predição de valores genéticos denominado BLUP - Melhor Preditor Linear não Viesado (*Best Linear Unbiased Prediction*) é um modelo misto, desenvolvido em 1949 e apresentado formalmente por Henderson (1973). Consiste em um procedimento que permite a predição dos valores genéticos de animais. Caracterizado como um

modelo misto, por tratar tanto dos efeitos fixos quanto dos aleatórios, o BLUP utiliza informações de todos os parentes, estabelecendo uma matriz de parentesco e possibilita a comparação de animais entre diferentes rebanhos. O BLUP faz a distinção entre os componentes genéticos e os fatores de ambiente, sendo capaz de trata-los de maneira individual para aumentar a acurácia das informações geradas acerca dos valores genéticos de um indivíduo. É capaz, também, de acordo com Mrode (2014), de calcular o coeficiente de parentesco entre indivíduos e atribuir pesos diferentes para suas características baseado em uma matriz de consanguinidade.

A identificação de animais geneticamente superiores depende da utilização de um método preciso para a seleção das melhores características e predição dos valores genéticos dos indivíduos candidatos à seleção. Entretanto, de acordo com Nogueira *et al.* (2003), a demanda computacional para esta análise é extremamente elevada e, por esta razão, o método BLUP só foi amplamente difundido a partir dos anos 90 quando, diante do avanço tecnológico daquela época, novas estruturas computacionais possibilitaram sua adoção em maior escala. Sua aceitação aumentou até ele se tornar o procedimento padrão de avaliação genética em bovinos de corte, bovinos de leite, ovinos e suínos. Nos dias atuais, a predição de características genéticas é um estudo amplamente difundido e movimenta pesquisadores em vários países. O BLUP segue sendo o procedimento mais utilizado para esse tipo de estimação, viabilizando o comparativo entre indivíduos de rebanhos diferentes e possibilitando, também, a predição dos valores genéticos de indivíduos que expressam seus fenótipos em períodos distintos. Desta forma, torna-se possível predizer a avaliação genética e os valores das próximas gerações, assim como identificar animais geneticamente superiores, independente do rebanho onde se encontrem, para assegurar a melhor seleção.

A resolução do BLUP consiste em uma série de etapas, capazes de minimizar a variância do erro de predição, estabelecer os efeitos fixos de um determinado atributo e os valores genéticos para cada indivíduo (JIANG, 2007). A metodologia dos modelos mistos é baseada em álgebra linear e diversas linguagens estatísticas podem ser utilizadas para a implementação computacional de um modelo. O exemplo a seguir, utiliza um conjunto de dados amostrais, representados pela Tabela 1, onde são apresentados os valores de ganho de peso pós-desmame, *Weaning Weight Gain* (WWG) de um grupo de animais. O objetivo do BLUP, neste caso, é estimar os efeitos do sexo e prever valores genéticos para todos os animais. Os desdobramentos da equação, assim como o código fonte do *script*, estão disponibilizados na seção de apêndices. A notação matricial que descreve este modelo

animal é mostrada pela Fórmula 2:

$$\gamma = X \beta + Z \alpha + \varepsilon \quad (2)$$

Onde:

- $\gamma$ : ( $n \times 1$ ) vetor de observações;
- $\beta$ : ( $p \times 1$ ) vetor dos efeitos fixos;
- $\alpha$ : ( $q \times 1$ ) vetor dos efeitos aleatórios;
- $\varepsilon$ : ( $n \times 1$ ) vetor de erros aleatórios;
- $X$ : matriz de incidência ( $n \times p$ ) que relaciona registros a efeitos fixos;
- $Z$ : matriz de incidência ( $n \times q$ ) que relaciona registros a efeitos aleatórios;
- $n$ : número de registros;
- $p$ : número de níveis para efeitos fixos;
- $q$ : número de níveis para efeitos aleatórios.

Desta forma,  $\gamma$  é o vetor que corresponde aos valores fenotípicos observados nos animais, ou seja, contém tantas linhas quanto forem o número de animais ( $n$ ) e seus valores são referentes às medições da característica observada. Neste caso, o ganho de peso ao desmame (WWG).  $\beta$  é um vetor, de dimensões ( $p \times 1$ ), que corresponde aos efeitos fixos, ou efeitos ambientais identificáveis, enquanto  $p$  é o número de níveis destes efeitos fixos. Neste caso, como o efeito correspondente é o sexo,  $\beta$  terá dois níveis, 0 e 1.  $\alpha$  corresponde ao vetor de efeitos aleatórios, ou genéticos dos animais, tem dimensão  $q \times 1$ , onde  $q$  é o número de níveis para estes efeitos. Neste caso terá 5 níveis.  $\varepsilon$  é um vetor, de dimensões  $n \times 1$ , correspondente aos efeitos ambientais não identificáveis, ou erros aleatórios. Neste caso,  $n$  corresponde à quantidade de indivíduos no conjunto, ou seja, todos os 8 animais. Finalmente,  $X$  e  $Z$  são matrizes de incidência, ou delineamento, e só contém valores 0 e 1.  $X$  tem dimensão  $n \times p$  e relaciona os registros a efeitos fixos, enquanto  $Z$  tem dimensão  $n \times q$  e relaciona seus registros aos valores genéticos dos animais. As MME sob o modelo animal, conforme Mrode (2014), podem ser formadas de acordo com a Fórmula 3:

$$\begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}a \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ Z'y \end{pmatrix} \quad (3)$$

Onde:  $A^{-1}$  representa a inversa da matriz de parentesco e  $a$  é o quociente de variâncias ambiental e genética, definido por  $a = \sigma_e^2/\sigma_a^2$ , respectivamente.

Tabela 1 – Ganho de peso, à desmama, para bezerros de corte.

<i>Animal</i>	<i>Sexo</i>	<i>Pai</i>	<i>Mãe</i>	<i>WWG (kg)</i>
4	M	1	-	4.5
5	F	3	2	2.9
6	F	1	2	3.9
7	M	4	5	3.5
8	M	3	6	5.0

Fonte: Mrode (2014)

A resolução do modelo misto por inversão da matriz de coeficientes gera um vetor, resultante do desdobramento da equação, de onde podem ser extraídos, nas primeiras linhas, os valores estimados para o efeito fixo do sexo e, logo a seguir, os valores estimados dos efeitos genéticos para cada indivíduo. O resultado, ilustrado pela Figura 1, indica que os bezerros machos possuem uma taxa de ganho de peso até o desmame maior do que bezerros do sexo feminino, o que é consistente com as médias brutas para machos e fêmeas. De acordo com os resultados obtidos, pode-se concluir que os animais machos têm ganho de peso médio ao desmame de 4.3584 kg e as fêmeas de 3.4044kg. Obteve-se também o valor genético de cada animal, a partir da terceira linha do vetor, onde ficou demonstrada a diferença de ganho de peso ao desmame entre cada animal em comparação com a média geral.

Como prova de conceito da implementação do BLUP, foi considerado o exemplo 3.1 de Mrode (2014), para analisar a característica de ganho de peso à desmama, assumindo que a variância genética é dada por  $\sigma^2 = 2$ . Os cálculos, desdobramento da equação e resultados finais estão disponíveis no Apêndice A.

## 2.4 Matriz de Parentesco

Uma das etapas para a resolução do BLUP é a inclusão das informações disponíveis acerca das relações de consanguinidade entre os indivíduos, por meio de uma matriz de parentesco. Jiang (2007) justifica a sua utilização pela possibilidade de conhecer o valor genético de um indivíduo a partir do valor correspondente dos seus parentes. A matriz de parentesco é capaz de estimar a proporção de genes idênticos entre indivíduos provenientes de um mesmo ascendente. Neste caso, quanto maior a quantidade de genes em comum, maior será o grau de parentesco e a semelhança de desempenho. A Tabela 2 mostra um conjunto de animais e seus ascendentes. Na primeira coluna estão enumerados todos os animais do conjunto, nas colunas 2 e 3 são relacionados os seus respectivos ascendentes, quando existentes.

Tabela 2 – Pedigree para bezerros de corte

<i>Animal</i>	<i>Pai</i>	<i>Mãe</i>
1	-	-
2	-	-
3	-	-
4	1	-
5	3	2
6	1	2
7	4	5
8	3	6

Fonte: Adaptada de Mrode (2014)

Figura 1 – Resultados do BLUP para ganho de peso à desmama em bovinos de corte



Fonte: Autor (2019)

De acordo com a Tabela 2, os animais 1, 2 e 3, não possuem ascendentes conhecidos. Mesmo assim, devem ser considerados na matriz de parentesco, por possuírem descendentes dentro do conjunto. A Tabela 3 apresenta a matriz de parentesco para este conjunto de 8 animais, conforme o exemplo adaptado de Mrode (2014), onde é possível verificar a análise de pedigree. Nesta matriz, a diagonal principal se caracteriza por apresentar os valores mais elevados, consequência da comparação feita a partir dos valores de um indivíduo com ele próprio. Para os demais campos da matriz, são esperados valores próximos de zero onde não existe relação de parentesco. O grau de parentesco, ou coeficiente de endogamia, entre pais e filhos é sempre 0.5, pois cada descendente direto recebe metade do seu conteúdo genético de cada um dos pais. Entretanto, este percentual pode variar muito entre irmãos, meios-irmãos, ou qualquer outro grau de parentesco. Na relação entre irmãos, estima-se encontrar 50% dos alelos comuns. Porém, o percentual pode variar entre 0 e 100%. Entre meios-irmãos, a estimativa média é de 25% de alelos comuns, considerando uma variação possível entre 0 e 50%

Tabela 3 – Matriz de consanguinidade entre os indivíduos

	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$	$\alpha 6$	$\alpha 7$	$\alpha 8$
$\alpha 1$	1	0	0	0.5	0	0.5	0.25	0.25
$\alpha 2$		1	0	0	0.5	0.5	0.25	0.25
$\alpha 3$			1	0	0.5	0	0.25	0.5
$\alpha 4$				1	0	0.25	0.5	0.125
$\alpha 5$					1	0.25	0.5	0.375
$\alpha 6$						1	0.25	0.5
$\alpha 7$							1	0.25
$\alpha 8$								1

Fonte: Autor (2019)

A montagem da matriz de parentesco, representada no sistema de equações pela variável  $A$ , tem seu ponto de partida na distribuição de valores, entre 0 e 1, referentes ao coeficiente de endogamia, onde 0 significa nenhuma correlação e 1 significa 100% de correlação entre os alelos. Na diagonal principal ( $a_{ij}$ ), onde  $i=j$ , como são comparados os valores do indivíduo com ele mesmo, sempre será atribuído 100% de endogamia, ou seja, 1. Em seguida, o valor atribuído para cada indivíduo ( $i$ ) é igual a  $(1+F_i)$ , onde  $F_i$  é o coeficiente de endogamia do indivíduo ( $i$ ). A matriz gerada é quadrada e os demais elementos, fora desta diagonal, são referentes ao coeficiente de parentesco entre os indivíduos. Cada valor está correlacionado ao grau de parentesco que o animal correspondente da linha tem

em relação ao animal correspondente da coluna.

Antes do preenchimento da matriz, devem ser seguidos alguns procedimentos de padronização dos dados, conforme os preceitos de Henderson (1976), que consistem em organizar o conjunto de animais, de forma que todos os ascendentes precedam suas pro-gênes e, codificá-los, em ordem numérica e exclusiva. Além disso, deve ser observada a existência de parentais conhecidos e aplicadas as regras apropriadas para cada caso:

- Quando  $s$  (*sire*) e  $d$  (*dam*), respectivamente pai e mãe do animal ( $i$ ), são conhecidos, aplica-se o seguinte conjunto de regras:
  - ◊  $a_{ii} = 1 + 0,5 (a_{sd})$ ;
  - ◊  $a_{ji} = a_{ij} = 0,5 (a_{js} + a_{jd}) ; j = [1:(i-1)]$ .
- Quando apenas um dos pais ( $s$ ) é conhecido e supostamente não possui grau de parentesco com a matriz, aplica-se:
  - ◊  $a_{ii} = 1$ ;
  - ◊  $a_{ji} = a_{ij} = 0,5 (a_{js}) ; j = [1:(i-1)]$ .
- Quando ambos os parentais são desconhecidos, aplica-se:
  - ◊  $a_{ii} = 1$ ;
  - ◊  $a_{ji} = a_{ij} = 0 ; j = [1:(i-1)]$ .

Onde:

- $i$ : corresponde ao indivíduo ( $i$ );
- $j$ : corresponde ao indivíduo ( $j$ );
- $s$ : corresponde ao pai, ou reprodutor (do inglês, *sire*);
- $d$ : corresponde à mãe, ou matriz (do inglês, *dam*);
- $a_{sd}$ : representa o coeficiente de parentesco entre ( $s$  e  $d$ );
- $a_{js} + a_{jd}$ : é a soma dos coeficientes de parentesco de ( $j$  e  $s$ ) e ( $j$  e  $d$ );
- $a_{js}$ : representa o coeficiente de parentesco entre ( $j$  e  $s$ ).

Considerando o pedigree da Tabela 2, obteve-se a representação matricial dada pela Tabela 3 mediante o seguinte conjunto de elementos:

$$a_{11} = 1;$$

$$a_{22} = 1;$$

$$a_{33} = 1;$$

$$a_{44} = 1 + 0 = 1;$$

$$a_{55} = 1 + 0 = 1;$$

$$a_{66} = 1 + 0 = 1;$$

$$a_{77} = 1 + 0 = 1;$$

$$a_{88} = 1 + 0 = 1;$$

$$a_{12} = 0;$$

$$a_{13} = 0;$$

$$a_{14} = 0,5 (a_{js}) = 0,5 (a_{11}) = 0,5;$$

$$a_{15} = 0,5 (a_{js} + a_{jd}) = 0,5 (a_{13} + a_{12}) = 0,5 (0) = 0;$$

$$a_{16} = 0,5 (a_{11} + a_{12}) = 0,5 (1 + 0) = 0,5;$$

$$a_{17} = 0,5 (a_{14} + a_{15}) = 0,5 (0,5 + 0) = 0,25;$$

$$a_{18} = 0,5 (a_{13} + a_{16}) = 0,5 (0 + 0,5) = 0,25$$

$$a_{23} = 0;$$

$$a_{24} = 0,5 (a_{21}) = 0;$$

$$a_{25} = 0,5 (a_{23} + a_{22}) = 0,5 (0 + 1) = 0,5;$$

$$a_{26} = 0,5 (a_{21} + a_{22}) = 0,5 (1) = 0,5;$$

$$a_{27} = 0,5 (a_{24} + a_{25}) = 0,5 (0,5) = 0,25;$$

$$a_{28} = 0,5 (a_{23} + a_{26}) = 0,5 (0,5) = 0,25;$$

$$a_{34} = 0,5 (a_{31}) = 0;$$

$$a_{35} = 0,5 (a_{33} + a_{32}) = 0,5;$$

$$a_{36} = 0,5 (a_{31} + a_{32}) = 0;$$

$$a_{37} = 0,5 (a_{34} + a_{35}) = 0,25;$$

$$a_{38} = 0,5 (a_{33} + a_{36}) = 0,5;$$

$$a_{45} = 0,5 (a_{43} + a_{42}) = 0;$$

$$a_{46} = 0,5 (a_{41} + a_{42}) = 0,25;$$

$$a_{47} = 0,5 (a_{44} + a_{45}) = 0,5 (0,25 + 0,25) = 0,25;$$

$$a_{48} = 0,5 (a_{43} + a_{46}) = 0,5 (0,25) = 0,125;$$

$$a_{56} = 0,5 (a_{51} + a_{52}) = 0,25;$$

$$a_{57} = 0,5 (a_{54} + a_{55}) = 0,5;$$

$$a_{58} = 0,5 (a_{53} + a_{56}) = 0,5 (0,25 + 0,25) = 0,375;$$

$$a_{67} = 0,5 (a_{64} + a_{65}) = 0,5 (0,25 + 0,25) = 0,25;$$

$$a_{68} = 0,5 (a_{63} + a_{66}) = 0,5;$$

$$a_{78} = 0,5 (a_{73} + a_{76}) = 0,5 (0,25 + 0,25) = 0,25;$$

O BLUP utiliza a inversa da matriz obtida pelos cálculos dos elementos acima. Para tanto, Henderson (1976) define um método para preenchimento da matriz quadrada ( $n \times n$ )  $A^{-1}$ , sem necessidade de calcular a inversão. Considerando a ordenação numérica e a regra de precedência dos ascendentes perante suas progênies, são aplicadas as seguintes regras:

- Quando  $s$  (*sire*) e  $d$  (*dam*), são conhecidos, deve ser adicionado:
  - ◊ +2 ao elemento  $a_{ii}$ ;
  - ◊ -1 aos elementos  $a_{is}$ ,  $a_{si}$ ,  $a_{id}$  e  $a_{di}$ ;
  - ◊ 0,5 aos elementos  $ass$ ,  $asd$ ,  $ads$  e  $add$ .
- Quando apenas um dos ascendentes for conhecido, deve ser adicionado:
  - ◊ +4/3 ao elemento  $a_{ii}$ ;
  - ◊ -2/3 aos elementos  $a_{is}$  e  $a_{si}$ ;
  - ◊ +1/3 ao elemento  $a_{ss}$ .
- Quando nenhum dos parentais do indivíduo ( $i$ ) for conhecido:
  - ◊ +1 ao elemento  $a_{ii}$ .

Onde:

- $i$  corresponde ao indivíduo ( $i$ );
- $s$  corresponde ao pai, ou reprodutor (do inglês, *sire*);
- $d$  corresponde à mãe, ou matriz (do inglês, *dam*).

A Tabela 4 mostra a formulação de novas equações, oriundas da aplicação destas regras, preenchendo os campos de uma matriz quadrada completamente zerada ( $A$ ). Em seguida, a Tabela 5 mostra a matriz invertida ( $A^{-1}$ ).

Tabela 4 – Procedimento para inversão da matriz (A)

	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$	$\alpha 6$	$\alpha 7$	$\alpha 8$
$\alpha 1$	$1 + \frac{1}{3} + \frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{2}{3}$	0	-1	0	0
$\alpha 2$		$1 + \frac{1}{2} + \frac{1}{2}$	$\frac{1}{2}$	0	-1	-1	0	0
$\alpha 3$			$1 + \frac{1}{2} + \frac{1}{2}$	0	-1	$\frac{1}{2}$	0	-1
$\alpha 4$				$\frac{4}{3} + \frac{1}{2}$	$\frac{1}{2}$	0	-1	0
$\alpha 5$					$2 + \frac{1}{2}$	0	-1	0
$\alpha 6$						$2 + \frac{1}{2}$	0	-1
$\alpha 7$							2	0
$\alpha 8$								2

Fonte: Autor (2019)

Tabela 5 – Matriz inversa de consanguinidade entre os indivíduos ( $A^{-1}$ )

	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$	$\alpha 6$	$\alpha 7$	$\alpha 8$
$\alpha 1$	1.833	0.500	0.000	0.667	0.000	1.000	0.000	0.000
$\alpha 2$	0.500	2.000	0.500	0.000	1.000	1.000	0.000	0.000
$\alpha 3$	0.000	0.500	2.000	0.000	1.000	0.500	0.000	1.000
$\alpha 4$	0.667	0.000	0.000	1.833	0.500	0.000	1.000	0.000
$\alpha 5$	0.000	1.000	1.000	0.500	2.500	0.000	1.000	0.000
$\alpha 6$	1.000	1.000	0.500	0.000	0.000	2.500	0.000	1.000
$\alpha 7$	0.000	0.000	0.000	1.000	1.000	0.000	2.000	0.000
$\alpha 8$	0.000	0.000	1.000	0.000	0.000	1.000	0.000	2.000

Fonte: Mrode (2014)

## 2.5 Modelos de predição robusta

Modelos de análise cujos resultados não são sensíveis à presença de *outliers* são denominados modelos robustos, dentro do contexto do melhoramento genético (CARDOSO *et al.*, 2009). Um *outlier* é uma observação, ou um subconjunto de observações, que extrapola a quantidade de desvios padrão definidos como aceitáveis, ao ponto de enviesar o resultado de uma análise, ou seja, é um dado extremo, cujo valor foge da normalidade e causa distorções no resultado de uma avaliação (SIROL *et al.*, 2007). Entretanto, dependendo do caso, esse valor extremo pode ser justamente o que se busca em um estudo (BARBOSA; PEREIRA; OLIVEIRA, 2018). Em pequenas amostras, como a Tabela 1,

a tarefa de encontrar valores extremos pode ser feita de forma visual, pelo pesquisador. Todavia, quando se trata de um grande conjunto de variáveis ou, quando a quantidade de observações extrapola a casa dos milhares, esta atividade se torna impraticável. Normalmente, os pesquisadores optam pela eliminação dos *outliers*, do conjunto de dados, como forma de preservar a normalidade na distribuição dos dados. De acordo com Costa *et al.* (2013), um valor inconsistente com as demais partes do conjunto é considerado disperso dentro de uma amostragem e precisa ser eliminado para que não exerça influência no resultado da análise, pois geraria distorção na média e desvio padrão de todo o conjunto de dados.

A eliminação de dados, apesar de ser uma prática comum, não é a única forma de tratamento para os *outliers*. Barbosa, Pereira e Oliveira (2018) consideram que as alternativas mais utilizadas para tratamento de valores extremos são as seguintes:

- **Exclusão:** consiste em eliminar as observações discrepantes da amostra de dados. A exclusão geralmente é aplicada quando o valor discrepante é considerado fruto de um possível erro na entrada dos dados e, não havendo possibilidade de releitura, ele precisa ser removido da amostra;
- **Análise individual:** esta abordagem é interessante quando o pesquisador busca investigar casos extremos, como animais com ganho de peso muito acima da média devido a tratamento diferenciado ou ganho negativo de peso, que pode ser indicativo de doenças, entre outros;
- **Métodos de clusterização:** Uma técnica de agrupamento que visa classificar os valores de acordo com a proximidade. Sendo capaz de aproximar o valor dos dados discrepantes dos seus vizinhos mais próximos. Reduzindo, dessa forma, a diferença através da atribuição de um novo valor aos *outliers*. É viabilizada por algoritmos de clusterização que fazem inferência de qual seria o melhor valor aproximado, baseado no comportamento das observações mais próximas ao dado *outlier*.

De acordo com (CARDOSO *et al.*, 2009), a estatística clássica, onde se incluem os modelos mistos como o BLUP, é sensível a dados extremos, devido ao fato que as estimativas dos parâmetros e valores genéticos podem ser deformadas por observações que sofreram influência de um ou mais fatores alheios ao modelo estatístico e diferem consideravelmente da média, podendo provocar anomalias nos resultados obtidos pelos modelos de análise. Não existindo uma regra para a censura de dados, a decisão fica por conta do pesquisador, que determina um limite para que as observações mais afastadas

da média sejam consideradas aproveitáveis, ou não, dentro do modelo estatístico (PIRES; BRANCO, 2007).

Programas de avaliação genética, como o PampaPlus (CARDOSO; LOPA; TEIXEIRA, 2016), consideram válidos os valores até o limite de desvios padrão estabelecido (geralmente, entre 3 e 4). São considerados *outliers*, os valores que extrapolarem esta determinada quantidade de desvios padrão. A quantidade de desvios padrão aceitáveis pode variar conforme o tamanho da amostra ou nível de precisão definidos pelo pesquisador (COSTA *et al.*, 2013). Atualmente, o PampaPlus considera valores até 3,5 desvios padrão e observações que extrapolem estes limites são desprezadas. Yokoo *et al.* (2013) considera que na análise de valores genéticos de bovinos de corte, nem todas as características analisadas resultam em uma distribuição normal, ou possuem *outliers*. E que, desta forma, a abordagem estatística clássica nem sempre é a mais adequada para estes casos. Neste sentido, a implementação de avaliações robustas, recomendada por (CARDOSO *et al.*, 2009) é capaz de obter estimativas mais fidedignas e confiáveis do valor genético dos indivíduos.

A predição robusta é um método que considera a análise de todo o conjunto de dados, reduzindo proporcionalmente o fator de influência das observações mais extremas, sem descartá-las. Este procedimento é possível devido à inclusão de uma variável de ponderação, que atribui um grau de relevância diferenciado para cada observação, conforme aumenta a distância para a média. Desta forma, apesar de ter seu fator de influência reduzido, nenhuma observação é descartada do modelo (CARDOSO; OLIVEIRA, 2015). O conjunto de dados da Tabela 1 foi considerado para apresentar o modelo robusto. Para simular uma entrada de dados incorreta foi introduzido um erro proposital, no valor genético do sétimo animal, transformando um dado legítimo em *outlier*, conforme a Tabela 6.

Tabela 6 – Dados de WWG com *outlier*

<i>Animal</i>	<i>Sexo</i>	<i>Pai</i>	<i>Mãe</i>	<i>WWG (kg)</i>
4	M	1	-	4.5
5	F	3	2	2.9
6	F	1	2	3.9
7	M	4	5	35
8	M	3	6	5.0

Fonte: Autor (2019)

Foi aplicado o BLUP, conforme demonstrado no apêndice A, e o resultado final sofreu uma drástica alteração, influenciado pelo valor do sétimo animal que elevou a média geral de ganho de peso ao desmame do grupo dos machos. Esta elevação da média fez com que o restante dos indivíduos deste mesmo grupo tivesse um desempenho negativo. A Tabela 7 demonstra a distorção dos resultados, após o cálculo do MME, ocasionada pela preservação de valores extremos nos dados de características animal. Na segunda e terceira colunas são apresentados os resultados obtidos após o cálculo do BLUP com o conjunto de valores original e alterado, respectivamente.

Tabela 7 – Predição de valores genéticos com o BLUP, considerando os valores extremos

	<i>Valor Real</i>	<i>Inferência com BLUP</i>
Média do grupo $\beta 1$	4.35850	14.16491
Média do grupo $\beta 2$	3.40443	2.50203
Valor genético de $\alpha 1$	0.09844	-0.70471
Valor genético de $\alpha 2$	-0.01877	1.12136
Valor genético de $\alpha 3$	-0.04108	-0.18889
Valor genético de $\alpha 4$	-0.00866	-0.69398
Valor genético de $\alpha 5$	-0.18573	2.06456
Valor genético de $\alpha 6$	0.17687	-0.26863
Valor genético de $\alpha 7$	-0.24945	4.71524
Valor genético de $\alpha 8$	0.18261	-2.01599

Fonte: Autor (2019)

Os resultados apresentados na Tabela 7 indicam que o cálculo de valores genéticos para o conjunto de animais sem a aplicação do modelo robusto, é sensível a *outliers* e teve o seu resultado distorcido por uma única observação com valor extremo. Neste exemplo, o cálculo da média considera a soma de todos os valores, dividido pelo número de indivíduos no grupo. Desta forma, o valor de  $\alpha 7$  faz com que a média de ganho diário de peso do grupo dos animais machos aumente, de 4,35kg para 14,1kg. A aplicação da matriz de ponderação teria um efeito significativo neste cenário, pois reduziria o fator de influência de  $\alpha 7$ , minimizando a distorção dos resultados.

## 2.6 Matriz de Ponderacao

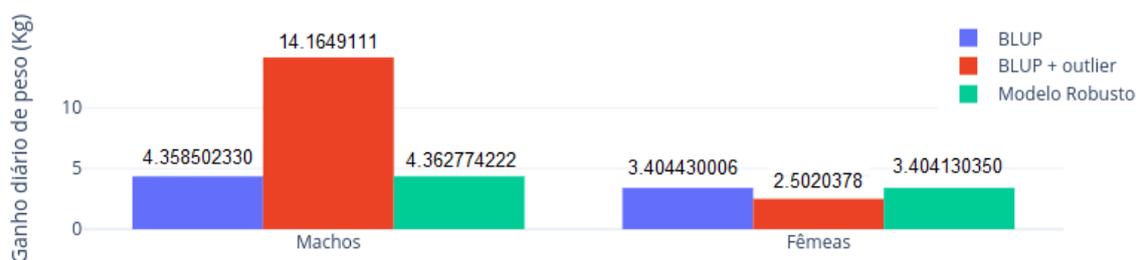
A matriz de ponderação é construída através da distribuição, em sua diagonal principal, de valores referentes ao ajuste do grau de relevância dos resultados estimados.

Possui dimensões  $n \times n$ , onde  $n$  é o número de indivíduos com fenótipos do conjunto, e todo o restante dos seus campos possui valor zerado. Segundo Cardoso *et al.* (2009) sua inclusão no sistema de equações do modelo misto possibilita minimizar o impacto causado, por observações com valores extremos, no resultado da análise. Neste sentido, adaptando a equação anterior (Fórmula 3) com a inclusão da matriz de ponderação ( $D$ ), teríamos a Fórmula 4:

$$\begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} X'DX & X'DZ \\ Z'DX & Z'DZ + A^{-1}a \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ Z'y \end{pmatrix} \quad (4)$$

Desta forma, resultados mais acurados terão os seus valores multiplicados por uma variável de ponderação próxima de 1, enquanto valores menos acurados deverão ser multiplicados por uma variável de ponderação mais próxima de zero. Levando em conta os resultados da Tabela 7, os resultados da estimação com o BLUP não possuem correlação alguma com o valor real, pois a média do grupo  $\beta_1$  foi elevada e isso enviesou os demais resultados da análise. Nesta situação, a predição dos valores genéticos considerando a matriz de ponderação é mais acurada, quando comparada com a predição feita com o BLUP, pois é capaz de minimizar os efeitos do dado, com valor extremo, sobre o grupo. A Figura 2 mostra uma comparação, acerca da predição dos valores genéticos, dos dados com *outlier* apresentados na Tabela 6. Os resultados estão divididos em dois grupos, de acordo com o sexo.

Figura 2 – Predição de valores genéticos em conjuntos com *outliers*



Fonte: Autor (2019)

Na primeira barra da Figura 2 são mostrados os valores reais, ou seja, valores que realmente correspondem ao ganho de peso diário dos animais e servem como parâmetro

para comparar a acurácia das predições a seguir. Na segunda barra é mostrado o resultado da predição com o BLUP e, na última barra, o resultado da predição considerando a matriz de ponderação.

O preenchimento da matriz de ponderação é um processo iterativo, dado pelo cálculo do erro padrão da predição, que deve ser refeito a cada iteração. Entretanto, a regra é sempre a mesma: preenchimento da diagonal principal com o valor da variável de ponderação e o restante dos campos zerados. A Tabela 8 mostra a alteração nos valores da matriz ( $D$ ) a cada iteração. A obtenção do valor, referente à variável ponderação ( $d$ ) é baseada na Fórmula 5:

$$d = \mu / (\mu + \varepsilon^2 / (\sigma * \nu)) \quad (5)$$

onde:

- $\mu$ : significa o número de observações consideradas no modelo;
- $\sigma$ : corresponde ao coeficiente de herdabilidade;
- $\nu$ : é relativo aos graus de liberdade da distribuição t;
- $\varepsilon$ : representa o valor do erro padrão da predição, e é obtido através da Fórmula 6:

$$\varepsilon = \gamma - X * R[1 : p] - Z * R[p + 1 : n] \quad (6)$$

onde:

- ◇  $\gamma$ : é o vetor de valores fenotípicos;
- ◇  $X$ : é a matriz de incidência referente aos efeitos fixos;
- ◇  $R$ : é o vetor de resultados;
- ◇  $R[1:p]$ : é um vetor com os valores genéticos dos grupos, ou efeitos fixos;
- ◇  $p$ : é a quantidade de níveis para efeitos fixos;
- ◇  $Z$ : é a matriz de incidência para efeitos aleatórios;
- ◇  $R[p+1:n]$ : é um vetor, referente aos valores genéticos individuais;
- ◇  $n$ : é a quantidade de indivíduos.

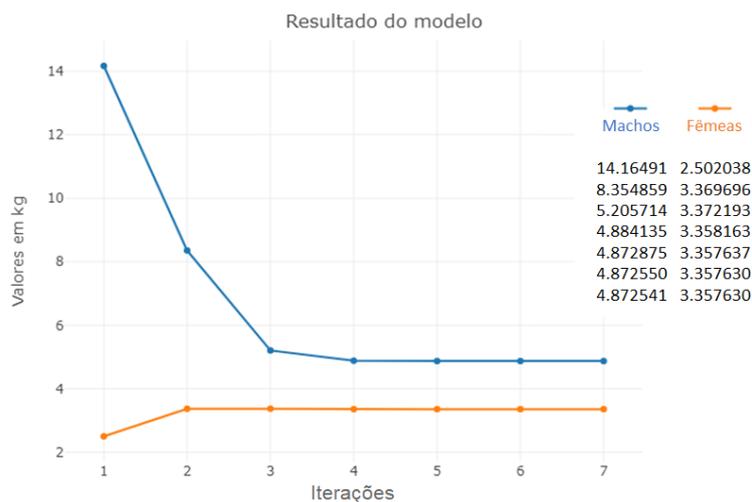
Conforme mostra a Tabela 8, os resultados tendem a convergir, após algumas iterações, e as variáveis de ponderação, neste caso, deixaram de promover alterações nas primeiras cinco casas decimais, o que determinou o encerramento do laço de repetição. Desta forma, foram obtidos os valores utilizados na equação que originou o resultado

Tabela 8 – Matriz de ponderação ( $D$ ) em cada iteração

<i>Iterações</i>	<i>Matriz de ponderação</i>
1 <sup>a</sup> iteração	$D = \begin{pmatrix} 0.09041845 & 0 & 0 & 0 & 0 \\ 0 & 0.742283 & 0 & 0 & 0 \\ 0 & 0 & 0.7422831 & 0 & 0 \\ 0 & 0 & 0 & 0.02986755 & 0 \\ 0 & 0 & 0 & 0 & 0.1353478 \end{pmatrix}$
2 <sup>a</sup> iteração	$D = \begin{pmatrix} 0.3456739 & 0 & 0 & 0 & 0 \\ 0 & 0.9744894 & 0 & 0 & 0 \\ 0 & 0 & 0.9744894 & 0 & 0 \\ 0 & 0 & 0 & 0.01131805 & 0 \\ 0 & 0 & 0 & 0 & 0.4334077 \end{pmatrix}$
3 <sup>a</sup> iteração	$D = \begin{pmatrix} 0.9407813 & 0 & 0 & 0 & 0 \\ 0 & 0.9808781 & 0 & 0 & 0 \\ 0 & 0 & 0.9808781 & 0 & 0 \\ 0 & 0 & 0 & 0.00895985 & 0 \\ 0 & 0 & 0 & 0 & 0.9944744 \end{pmatrix}$
4 <sup>a</sup> iteração	$D = \begin{pmatrix} 0.9855943 & 0 & 0 & 0 & 0 \\ 0 & 0.9812126 & 0 & 0 & 0 \\ 0 & 0 & 0.9812126 & 0 & 0 \\ 0 & 0 & 0 & 0.008750956 & 0 \\ 0 & 0 & 0 & 0 & 0.9996586 \end{pmatrix}$
5 <sup>a</sup> iteração	$D = \begin{pmatrix} 0.9866852 & 0 & 0 & 0 & 0 \\ 0 & 0.9812512 & 0 & 0 & 0 \\ 0 & 0 & 0.9812512 & 0 & 0 \\ 0 & 0 & 0 & 0.008743005 & 0 \\ 0 & 0 & 0 & 0 & 0.9995456 \end{pmatrix}$
6 <sup>a</sup> iteração	$D = \begin{pmatrix} 0.9867146 & 0 & 0 & 0 & 0 \\ 0 & 0.9812531 & 0 & 0 & 0 \\ 0 & 0 & 0.9812531 & 0 & 0 \\ 0 & 0 & 0 & 0.008742768 & 0 \\ 0 & 0 & 0 & 0 & 0.999542 \end{pmatrix}$
7 <sup>a</sup> iteração	$D = \begin{pmatrix} 0.9867154 & 0 & 0 & 0 & 0 \\ 0 & 0.9812531 & 0 & 0 & 0 \\ 0 & 0 & 0.9812531 & 0 & 0 \\ 0 & 0 & 0 & 0.008742761 & 0 \\ 0 & 0 & 0 & 0 & 0.9995419 \end{pmatrix}$

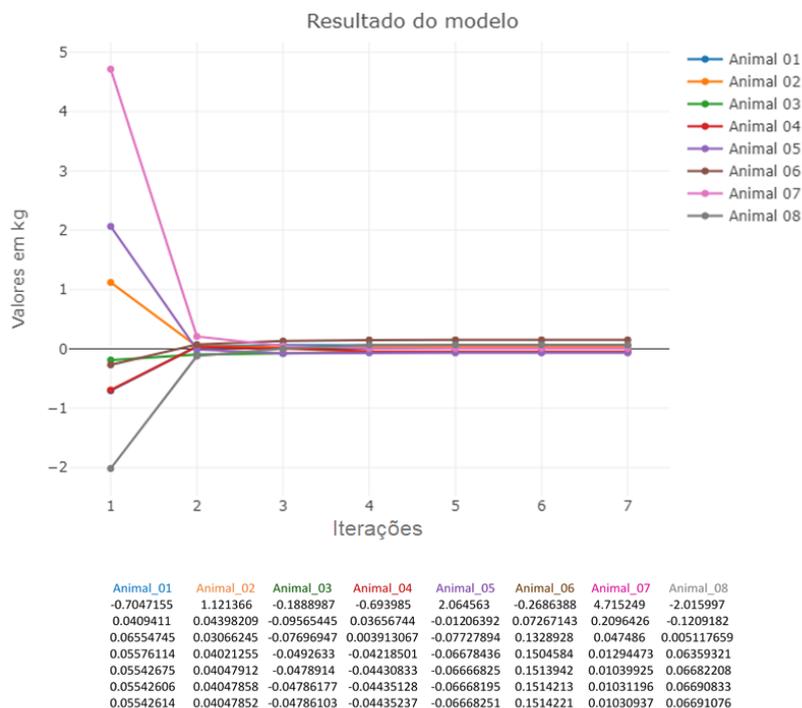
apresentado nas últimas barras (modelo robusto) da Figura 2. As Figuras 3 e 4 apresentam a variação dos resultados em cada um das iterações demonstrando a convergência dos valores, tanto para os grupos de machos e fêmeas, quanto para os valores genéticos individuais.

Figura 3 – Oscilação das médias dos valores genéticos, para machos e fêmeas, a cada iteração.



Fonte: Autor (2019)

Figura 4 – Oscilação dos valores genéticos individuais, a cada iteração.



Fonte: Autor (2019)

## 2.7 Trabalhos Correlatos

Durante a etapa de revisão de literatura, foram analisados alguns periódicos especializados, em busca de trabalhos relacionados com os temas pesquisados e que pudessem embasar, ou contribuir com a pesquisa. A busca levou em consideração a relevância do periódico publicado e a correspondência, de termos chaves, com os da presente pesquisa. Alguns destes trabalhos serão relacionados a seguir:

- No artigo de Copt e Victoria-Feser (2006) é descrito o processo de criação de um estimador robusto e proposta a utilização de um teste robusto para hipóteses multivariadas. Os autores consideram que o estimador de máxima verossimilhança (MLE) ou o MLE residual (REML) são geralmente escolhidos para estimar os parâmetros com base na suposição de normalidade multivariada exata. Mas que, no entanto, estes estimadores são sensíveis a desvios da normalidade multivariada fazendo com que, na prática, um conjunto de dados, ou mesmo um único dado isolado, com valor extremo, possa impulsionar o valor das estimativas por conta própria. São discutidas as vantagens deste estimador em comparação com os seguintes estimadores: (RICHARDSON; WELSH, 1995), (RICHARDSON, 1997), (WELSH; RICHARDSON, 1997) e (GILL, 2000). O material utilizado consiste em uma base de dados referente a 10mil amostras e a metodologia utilizada foi o cálculo do REML e do *Constrained Translated Biweight S-estimator* (CTBS), que é o modelo proposto, seguidos dos erros padrão correspondentes para os efeitos fixos. São apresentados gráficos de dispersão das distâncias de Mahalanobis estimadas usando os dois modelos e, as conclusões apresentadas sugerem que os resultados obtidos com o CTBS são mais acurados.
- O trabalho publicado por Copt e Heritier (2007) assume que os modelos mistos são geralmente ajustados por meio de técnicas de máxima verossimilhança restritas e que estas técnicas dependem da normalidade o modelo. Tal dependência é identificada como uma fraqueza, capaz de induzir a interpretações erradas. Diante disso, os autores justificam a utilização do modelo desenvolvido por Copt e Victoria-Feser (2006) como padrão para cálculos robustos de estimação e têm, como objetivo, desenvolver um novo modelo para calcular um teste do tipo razão de verossimilhança, uma alternativa robusta direta ao teste F. Os resultados

apresentados dão conta que a análise feita em um conjunto de dados reais indicou a vantagem da nova abordagem na presença de observações periféricas. Entretanto, os autores salientam que o modelo ainda não está concluído e, apesar dos resultados iniciais satisfatórios, ainda precisa ser aprimorado para que possa complementar o CTBS. A principal limitação é o fato que todos os estimadores e os testes relacionados foram desenvolvidos para projetos balanceados, ou seja, dados em que o mesmo número de medições é registrado por observação.

- Jamrozik, Fatehi e Schaeffer (2007) abordaram a aplicação de procedimentos robustos para predição de valores genéticos em um modelo de regressão aleatória com múltiplas características. A pesquisa considerou registros de um conjunto de 63.346 vacas Jersey canadenses, paridas entre 1990 e 2005, cujos valores diários de rendimento de leite, gordura e proteínas foram comparados, com o objetivo de classificar e ranquear os animais. A metodologia aplicada consiste na comparação de diferentes modelos de avaliação cujos critérios considerados foram: as estatísticas residuais, inspeção das distribuições de observações atípicas e a comparação de classificações de animais. Como resultado, os autores inferiram que a utilização de um método robusto foi capaz de reduzir a influência de *outliers*. Para tanto, os resultados obtidos pelos modelos robustos foram comparados com o BLUP.
- O trabalho de Cardoso *et al.* (2009) aborda a construção hierárquica do modelo animal sob o efoque bayesiano. Os autores descrevem e ilustram a aplicação de modelos bayesianos capazes de identificar *outliers* e atenuar a sua contribuição para a composição da média geral dos grupos. Estes modelos consideram o conhecimento obtido "a priori", assim como a informação contida nos dados para implementar a inferência através de métodos *Markov chain Monte Carlo* - Monte Carlo via Cadeias de Markov (MCMC). A metodologia proposta considera a introdução de uma variável que limita a contribuição de cada observação através da associação com valores proporcionalmente próximos de zero, fazendo com que observações com valores extremos contribuam menos para a estimativa dos parâmetros. A conclusão apresentada é que a utilização de modelos robustos é apropriada para identificar e mitigar o efeito de observações com valores extremos, na predição dos parâmetros do conjunto de dados observado. Salientam, entretanto,

que a intensa demanda computacional das análises baseadas no método exposto é um fator que limita a utilização em grande escala nos programas de avaliação genética.

- O trabalho de Gianola *et al.* (2018) trata da predição de características complexas, baseadas em informações de DNA, e apresenta alternativas robustas capazes de superar as limitações do método mais utilizado para estimar tais características na agricultura por genoma, o *Genomic Best Linear Unbiased Prediction* (GBLUP). O GBLUP é uma derivação do BLUP, onde a matriz de consanguinidade ( $A$ ) é substituída pela matriz de relação genômica ( $G$ ) que é baseada em marcadores de similaridades genômicas. Os autores apontam que uma potencial desvantagem do GBLUP, assim como acontece com o BLUP, é que os métodos de regressão linear sob as hipóteses de Gauss são sensíveis a *outliers*. Diante disso, o objetivo do trabalho é apresentar a implementação de alternativas robustas utilizando um modelo linear com distribuições  $t$  ou Laplace residuais, em vez de gaussianas, e avaliar cada um dos métodos conforme seus resultados. A metodologia utilizada consiste na combinação do método bayesiano *maximum a posteriori prediction* (MAP) com distribuições  $t$  ou Laplace denominadas TMAP e LMAP, respectivamente, para incorporar distribuições residuais de cauda grossa, mantendo uma estrutura computacional do tipo BLUP. O experimento utilizou 3 conjuntos de dados, com características distintas, para avaliar se TMAP ou o LMAP poderiam fornecer uma melhor capacidade preditiva do que o GBLUP. O primeiro conjunto, com dados de gado leiteiro, é representativo da aprendizagem observacional e consiste em 1.264 bovinos suíços da raça *Brown Brown* de 85 rebanhos pertencentes a criadores privados e é representativo da aprendizagem observacional. O segundo conjunto de dados foi baixado do pacote R *Synbreed* (WIMMER *et al.*, 2012) e aborda a expressão gênica de uma pequena planta, denominada *Arabidopsis thaliana*. Consiste em um conjunto de 199 acessos genotipados e é representativo de uma situação em que as condições experimentais são cuidadosamente controladas, como é frequentemente o caso dos organismos modelo. O terceiro conjunto, considera a existência de 599 linhagens de trigo, com 129 marcadores que denotam a presença ou ausência de um determinado alelo. Também disponível no R, através do pacote *BGLR* (PÉREZ; CAMPOS, 2014), este conjunto simula condições agronômicas com a possibilidade de organizar parcelas em campos e de randomização. De

maneira geral, para os 3 conjuntos de dados, ou autores concluíram que os modelos com distribuições residuais de  $t$  ou Laplace podem frequentemente fornecer previsões mais acuradas que o BLUP. Quanto ao GBLUP, os autores consideraram que os modelos robustos foram superiores quando utilizados o primeiro e segundo conjuntos de dados.

- O *Brazilian Breeding Values* (BRBV) é uma ferramenta em fase final de desenvolvimento, no âmbito do projeto MaxiDep (desenvolvimento e implementação de metodologias genético-estatísticas para avaliações em bovinos de corte), que é integrante do arranjo Maxibife, um dos maiores projetos de integração de conhecimento conduzidos pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), que atualmente conta com 27 organizações, distribuídas entre 10 unidades de pesquisa da EMBRAPA, empresas de desenvolvimento pecuário e universidades (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2017). Seu desenvolvimento teve como motivação a necessidade de oferecer suporte computacional para os programas de melhoramento animal, possibilitando a avaliação de populações de animais de grande porte. A versão atual é escrita em linguagem de programação *python*, utiliza uma pilha de pacotes de software para computação científica (*NumPy*, *SciPy* e *PyTables*) e implementa um solucionador BLUP para equações de modelo misto. Sua utilização na EMBRAPA justifica a escolha, em detrimento a outras possíveis soluções, para validação dos resultados obtidos com os modelos desenvolvidos ao longo do trabalho, mediante comparação de resultados entre as duas soluções, disponibilizadas no capítulo 4. Observou-se, após a implementação e testes, a possibilidade de contribuir com o BRBV, no sentido de reportar a demora no tempo de processamento do cálculo de acurácias, conforme demonstrado na seção 4.5. Considerou-se, também, o fato que até a presente versão, o BRBV não calcula a matriz de ponderação ( $D$ ). Esta limitação motivou o emprego de esforços para desenvolver um algoritmo capaz de contribuir para o ajuste de *outliers*.

Neste capítulo foram abordados os conceitos gerais que envolvem o melhoramento animal e a metodologia empregada para a predição dos valores genéticos, que são utilizados como parâmetros, capazes de determinar a seleção dos melhores indivíduos para acasalamento. Os exemplos utilizados são baseados no conjunto de dados disponibilizado por

Mrode (2014), onde o ganho de peso ao desmame (WWG) é utilizado como característica de avaliação, tendo o sexo como efeito fixo. Foram apresentados os principais conceitos acerca dos modelos robustos, sua aplicação e o cálculo da variável de ponderação. Por último, são apresentados alguns trabalhos correlatos, com destaque para o BRBV, que foi implementado e utilizado como prova de conceito. No próximo capítulo, será apresentada a metodologia utilizada para a predição de valores genéticos e identificação de *outliers* em dados pertencentes à EMBRAPA, onde foram analisadas duas características: a área de olho de lombo (AOL) e a espessura de gordura (EG). Ambas considerando os grupos contemporâneos (GC) como efeito fixo.

### 3 METODOLOGIA

Este capítulo descreve os procedimentos metodológicos que classificam a pesquisa, com base nos conceitos de caracterização definidos por Wazlawick (2017). São apresentadas também as etapas de elaboração, detalhando os procedimentos, material e métodos que foram utilizados durante o processo de pesquisa.

#### 3.1 Caracterização da pesquisa

- Quanto à natureza: é uma pesquisa aplicada, pois é capaz de gerar produtos ou processos e tem finalidades imediatas, utilizando conhecimentos prévios ou de pesquisas básicas para resolver problemas relacionados a aplicações concretas, envolvendo interesses locais. Neste caso, o produto em questão é um algoritmo capaz de solucionar o problema da análise de dados extremos.
- Quanto aos objetivos: Considera-se que a presente pesquisa é do tipo explicativa, pois interfere nos dados e interpreta os fenômenos observados. Segundo Gil (2002), uma pesquisa é explicativa quando aprofunda o conhecimento da realidade, explicando as causas através da análise, classificação e interpretação dos fenômenos.
- Quanto à abordagem: Wazlawick (2017) classifica, como quantitativas, aquelas pesquisas que são baseadas na quantificação dos dados coletados, explorando a geração de tabelas, gráficos e resultados baseados em números, para análise de opiniões e informações obtidas durante a pesquisa. O presente trabalho utilizou técnicas de estatística e recursos computacionais, procurando evidenciar o conhecimento descoberto através de números. Desta forma, o trabalho é enquadrado como quantitativo.
- Quanto aos procedimentos: Uma pesquisa científica é considerada experimental, de acordo com Wazlawick (2017), quando os procedimentos executados pelo pesquisador são capazes de provocar, sistematicamente, alterações no ambiente pesquisado, com o objetivo de produzir novos resultados a partir da intervenção proporcionada. Para Gil (2002), a pesquisa experimental ocorre quando existe delimitação do objeto de estudo e da seleção das variáveis que seriam capazes de influenciá-lo, associados com a definição das formas de controle e de observações que a variável produz no objeto. Nesta pesquisa, o objeto de estudo são os valores extremos,

através dos quais pretende-se executar procedimentos capazes de produzir novos resultados. Assim, entende-se que a pesquisa é do tipo experimental.

### 3.2 Material e métodos

Para o presente estudo foram consideradas informações de bovinos da raça Brangus, disponibilizadas pela EMBRAPA Pecuária Sul, pertencentes ao seu programa de avaliação genética, o PampaPlus. Programa este que tem como finalidade calcular os valores genéticos e índices, para designar as vantagens obtidas em usar determinado reprodutor ou matriz. A coleta dos dados, disponibilizados para esta pesquisa, segue as normas do programa PampaPlus (CARDOSO; LOPA; TEIXEIRA, 2016), que recomenda a participação de um especialista, indicado pela EMBRAPA, para realizar as aferições, medições e avaliações visuais.

Dois conjuntos de dados foram disponibilizados, pela EMBRAPA Pecuária Sul, para análise de características, com definições pré-estabelecidas de valores para variância genética (17,41) e variância residual (30,21): o primeiro, continha informações de 1387 bovinos, da raça Brangus, nascidos entre os anos de 1990 a 2017, onde foram feitas análises da Área de Olho de lombo (AOL) ao sobreano, considerando os Grupos Contemporâneos (GC) como efeito fixo. O segundo conjunto de dados foi uma versão atualizada do primeiro, obtida após as análises de AOL, contendo informações de 1552 animais, nascidos entre 1990 e 2018, onde foram feitas análises acerca da espessura de gordura (EG) tendo, também, os grupos contemporâneos como efeito fixo. As medidas de AOL e EG são aferidas através de exame de ultrassonografia, descrito em Rosa (2018). Os GC, de acordo com o Cardoso, Lopa e Teixeira (2016) são compostos por: (sexo + lote + ano de nascimento).

A análise descritiva dos dados utilizados está apresentada na Tabela 9, onde podem ser verificados maiores detalhes acerca das bases de dados utilizadas para cada uma das características. Os valores para AOL estão expressos em centímetros quadrados ( $cm^2$ ), que é a unidade de medida padrão desta característica. O mesmo vale para EG, que está expressa em milímetros (mm). Foi considerado o teste de normalidade de Shapiro Wilk, para calcular o valor de  $p$ , assumindo que um valor inferior a 0,05 indica que os dados não possuem uma distribuição normal (SHAPIRO; WILK, 1965). O desvio padrão indica a dispersão dos dados, dentro de uma amostra, com relação à média e foi obtido calculando a raiz quadrada da variância. Variância, que é obtida mediante a subtração de cada uma

das medidas individuais pela média e elevados os resultados ao quadrado sendo, em seguida, somados todos resultados e dividido o produto pelo tamanho da amostra menos um. O erro padrão é uma medida de variação de uma média amostral em relação à média da população e foi obtido através da divisão do desvio padrão pela raiz quadrada do tamanho amostral. Para ilustrar esta distribuição, a Figura 5 mostra a representação dos valores de AOL e EG através de histogramas individuais. A definição de *outliers* obedeceu o padrão definido em Cardoso, Lopa e Teixeira (2016), que é de 3,5 desvios padrão. Desta forma, foram identificadas 3 observações que extrapolaram os limites para a AOL e 9 observações para EG.

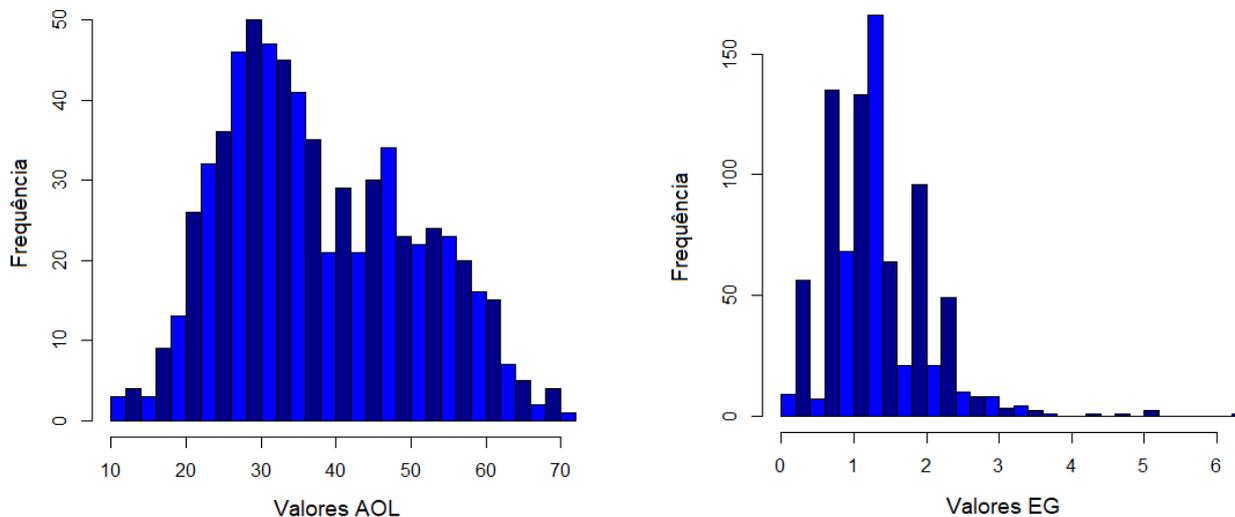
Tabela 9 – Análise descritiva dos dados

Característica	Área de Olho de Lombo	Espessura de Gordura
Unidade de medida	$cm^2$	mm
Quantidade de animais	1387	1552
Ano Inicial	1990	1990
Ano Final	2017	2018
Animais com fenótipos	687	866
Grupos Contemporâneos	32	52
Média da característica	37,83	13,8
Mediana da característica	35,40	14,0
Desvio Padrão da característica	12,6	6,88
Erro padrão	0,48	0,23
Valor máximo da característica	70,3	64,0
Valor mínimo da característica	11,5	0,0
Teste de normalidade	$p=3.416e-10$	$p=2.2e-16$

Fonte: Autor (2019)

O algoritmo desenvolvido é capaz de receber os dados fenotípicos e de pedigree, estimar os valores genéticos de acordo com o modelo animal e avaliar os resultados quanto à presença de *outliers*. Além disso, é capaz de calcular as variâncias e acurácias da predição. Como prova de conceito, com a finalidade de validar os resultados obtidos, foram replicadas as análises, com os mesmos conjuntos de dados e parâmetros no sistema BRBV, que é utilizado pela EMBRAPA, e aplicados testes para confirmar a correlação entre os dois resultados. A caracterização de quais dados são, ou não, *outliers*, foi ajustada de acordo com as definições da EMBRAPA, mas pode ser redefinida, de acordo com a necessidade do pesquisador. O fluxograma de execução do algoritmo é representado pela Figura 6 e mostra as etapas, que serão detalhadas a seguir, para: leitura e interpretação dos dados, cálculo das equações de modelos mistos, análise e mitigação de valores extremos.

Figura 5 – Histograma dos valores fenotípicos para as características AOL e EG.

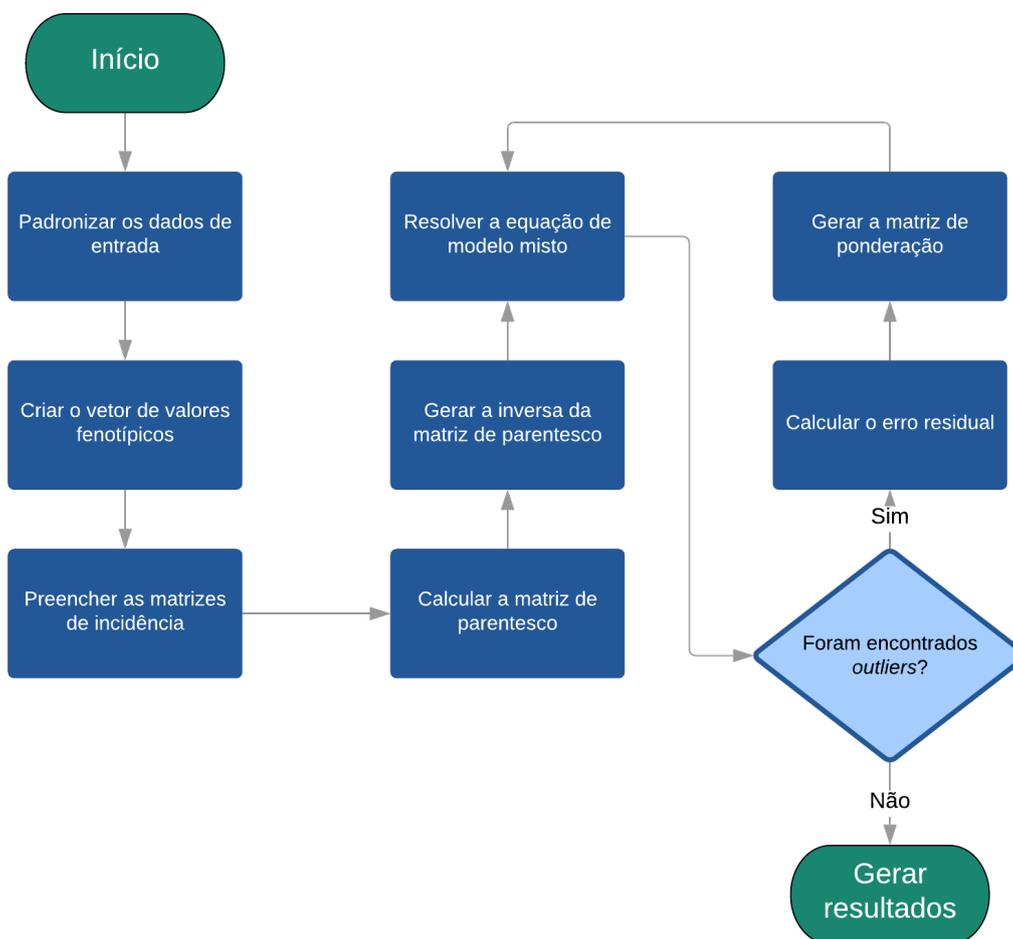


Fonte: Autor (2019)

Para adaptar os dados de entrada, prever os valores genéticos e mitigar as deformidades causadas pelos *outliers*, foi utilizado o *software* R (TEAM *et al.*, 2013), na versão 3.5.2, complementado pela biblioteca *pedigreemm* (VAZQUEZ *et al.*, 2010) versão 0.3-3, que possibilitou o cálculo da matriz de parentesco. Para a geração dos gráficos utilizados neste trabalho, foram utilizadas as bibliotecas *plotly* (SIEVERT *et al.*, 2017), na versão 4.9.0 e *ggplot2* (WICKHAM, 2011), na versão 3.2.1. Para a execução dos testes e análises de desempenho foi utilizado um servidor, com o sistema operacional Ubuntu versão 16.04.5, com 8gb de memória RAM e processador octacore de 64bits Intel(R) Xeon(R) CPU E7- 4830 @ 2.13GHz.

O procedimento adotado foi a aplicação do BLUP sobre o modelo animal, seguido do cálculo do erro padrão da predição e reajuste das variáveis, alocadas em uma matriz de ponderação, capazes de reduzir ou preservar a significância de cada observação dentro do conjunto. Este processo foi repetido, até que os valores convergissem com o resultado anterior. Foram considerados convergentes, os valores com diferença menor que duas casas decimais. A validação dos resultados das predições do BLUP obtidas com o R foi feita através do processamento dos mesmos dados no sistema BRBV versão 0.9.0. A correlação entre os resultados obtidos nos dois modelos foi de 100% e está detalhada na seção 4.3. Desta forma, considerando a equivalência dos resultados produzidos, com os resultados do sistema desenvolvido pela EMBRAPA, pode-se dizer que são corretas as predições dos valores genéticos obtidas pelo modelo desenvolvido.

Figura 6 – Fluxograma das etapas do processo de cálculo dos valores genéticos e mitigação dos efeitos causados por dados extremos.



Fonte: Autor (2019)

A padronização dos dados de entrada obedeceu às convenções de Henderson (1976), o que possibilitou a montagem das matrizes e o cálculo do modelo animal. Para tanto, foram filtrados somente os campos necessários para que os dados recebidos da EMBRAPA fossem importados com o seguinte layout:

- Arquivo de pedigree:

Primeira linha: cabeçalho do arquivo determinando a ordenação dos seguintes campos, separados por vírgula: *Animal, id\_animal, Pai, id\_pai, Mae, id\_mae*;

Corpo do arquivo: Valores correspondentes à nomenclatura e identificação exclusiva dos indivíduos e seu relacionamento parental, sendo aceitos apenas caracteres numéricos nos campos referentes à identificação;

- Arquivo de dados:

Primeira linha: cabeçalho do arquivo determinando a ordenação dos seguintes campos, separados por vírgula: *Animal*, *id\_animal*, *EF\_1*, *anonasc*, *Caract\_1*;

Corpo do arquivo: Valores correspondentes aos campos, onde:

- ◇ *Animal*: nomenclatura original do indivíduo;
- ◇ *Id\_animal*: identificação numérica e exclusiva, atribuída ao indivíduo;
- ◇ *EF\_1*: Efeito Fixo considerado;
- ◇ *Anonasc*: correspondente ao ano de nascimento; e
- ◇ *Caract\_1*: Característica avaliada.

O vetor de valores fenotípicos, conforme descreve a metodologia na seção 2.3, foi obtido diretamente do arquivo de dados. A criação das matrizes de incidência seguiu a metodologia descrita na seção 2.3 e seu processo está detalhado no apêndice A. O cálculo e a inversão da matriz de parentesco obedeceram o procedimento metodológico descrito na seção 2.4. O cálculo da MME foi através do modelo animal, cuja equação e a referida notação estão descritas na seção 2.3 e o detalhamento do processo está no apêndice A. O método para calcular o erro residual, a variável de ponderação e a resolução do modelo misto, considerando a matriz de ponderação, estão detalhados na seção 2.6 e, o cálculo das acurácias da predição está detalhado na seção 4.4.

Neste capítulo foram abordadas as classificações metodológicas, que caracterizam o presente estudo quanto à sua natureza, objetivos, abordagem e procedimentos. Além disso, foram apresentados os materiais, métodos e o detalhamento das etapas do processo, utilizados para a obtenção dos resultados que serão apresentados no capítulo 4.

## 4 RESULTADOS E DISCUSSÃO

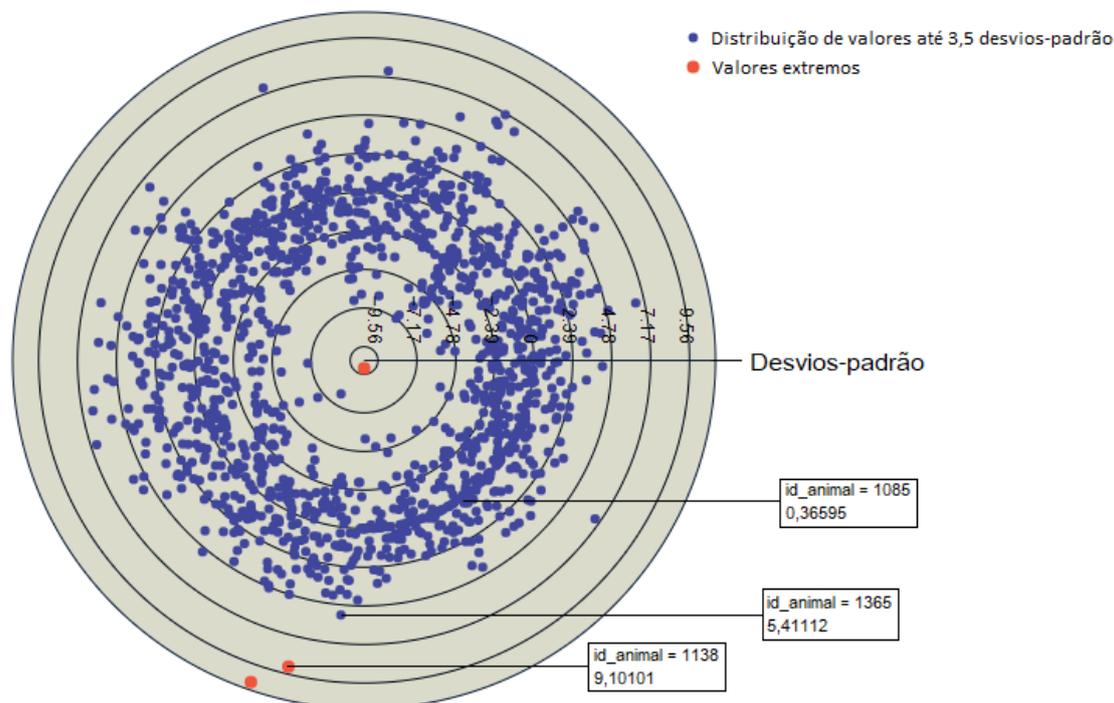
Neste capítulo serão apresentados os resultados obtidos com a aplicação do modelo robusto. Foram estimados os valores genéticos para duas características, área de olho de lombo (AOL) e espessura de gordura (EG). A validação dos cálculos das MME é apresentada através de uma comparação com o sistema BRBV. São apresentadas as acurácias e as correlações entre os resultados. Ao final do capítulo, são discutidas questões referentes ao desempenho computacional.

### 4.1 Predição dos valores genéticos da área de olho de lombo (AOL)

O cálculo da MME, para predição dos valores genéticos, apresentou resultados que determinaram a existência de *outliers* no conjunto. A Figura 7 mostra a distribuição dos valores genéticos estimados, onde cada ponto corresponde a um indivíduo e cada linha delimita um desvio padrão (DP). Nesta distribuição, a maioria dos elementos ficou concentrada entre a linha principal, cujo valor é zero, e as primeiras linhas concêntricas subsequentes, que representam a diferença de um desvio padrão, para mais ou menos. A quantidade de pontos entre as linhas concêntricas diminuiu, na medida em que se afastam da linha principal. A diminuição é gradativa, até restarem apenas três pontos distribuídos em regiões que extrapolam os limites de 3,5 desvios padrão.

Para explicar a distribuição dos valores da Figura 7 foram selecionados, de maneira intencional, 3 pontos em situações distintas: O primeiro, referente ao animal com número de identificação 1085, está localizado em uma região com a maior aglomeração de pontos, próxima à linha principal, isso significa que o dado está próximo do ajuste perfeito do modelo, que seria exatamente sobre a linha principal. Este indivíduo possui valor genético estimado de +0,36 e um desvio padrão de +0,15; O segundo ponto destacado é referente ao animal com id = 1365. É um dado com valor moderadamente extremo, ou seja, está afastado da linha principal, mas não o suficiente para extrapolar o limite de desvios padrão e ser considerado *outlier*. O valor genético estimado para este animal é de +5,41 e o desvio padrão calculado é de +2,26, em um cenário onde a proximidade do 0 é considerado o valor ótimo, ou desejável, e até 3,5 DP são aceitáveis. Por último, conforme destacado na Figura 7, o animal com id = 1138 possui valor genético estimado de +9,10, o que corresponde a +3,80 DP. Este é considerado um valor extremo e potencialmente capaz de enviesar o resultado da análise. Outros 2 pontos, em situação semelhante,

Figura 7 – Dispersão dos valores genéticos e desvios padrão após o cálculo da MME.



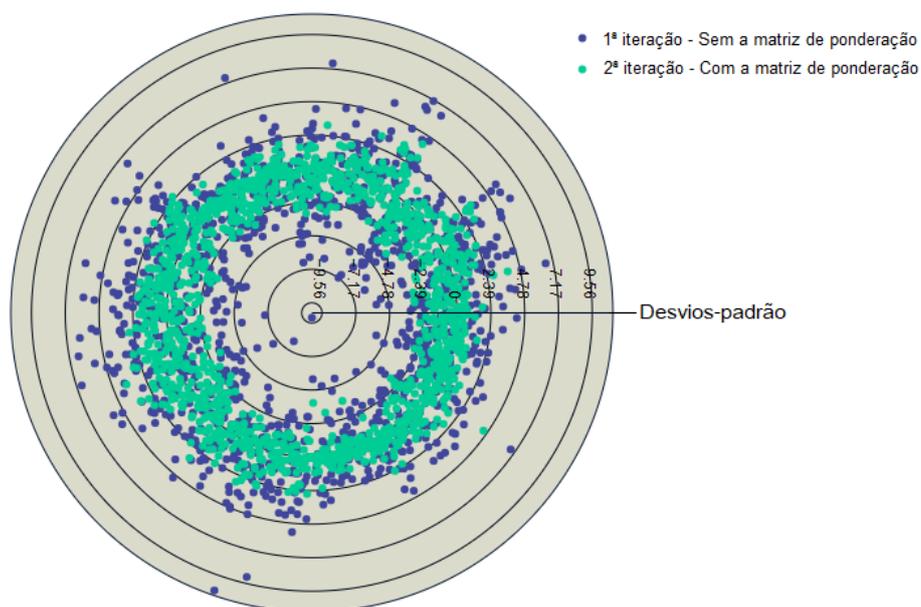
Fonte: Autor (2019)

estão caracterizados pela cor vermelha na Figura 7 e também são considerados *outliers*, pois extrapolaram o limite de 3,5 desvios padrão, para mais ou menos.

Dada a existência de *outliers*, uma nova execução do algoritmo refaz o cálculo da MME considerando, para tanto, a inclusão da matriz  $D$  na equação. Este procedimento atribui um grau de relevância, compreendido entre 0 e 1, para cada um dos elementos. Desta forma, um valor muito próximo da linha principal teve um multiplicador próximo de 1, preservando o seu potencial de relevância. Entretanto, valores mais afastados da linha principal tiveram os seus multiplicadores proporcionalmente mais próximos de 0. Como consequência, os valores genéticos de todo o conjunto tendem a aproximar-se da linha central do gráfico. A Figura 8 mostra, sobrepostos aos valores estimados na primeira execução, a distribuição dos valores genéticos após nova execução do algoritmo e recálculo da MME com a inclusão da matriz de ponderação.

A concentração de pontos em torno da linha principal aumentou após a execução da segunda iteração, onde foi recalculada a equação com o acréscimo da variável de ponderação. A Figura 8 demonstra que apenas uma nova execução do algoritmo foi suficiente para que não houvessem valores extremos na distribuição dos dados. Os 3 valores identificados como *outliers* e seus respectivos ajustes são apresentados na Tabela 10.

Figura 8 – Dispersão dos valores genéticos recalculados com a matriz de ponderação.



Fonte: Autor (2019)

Tabela 10 – Valores genéticos de *outliers* estimados sem matriz e com matriz de ponderação.

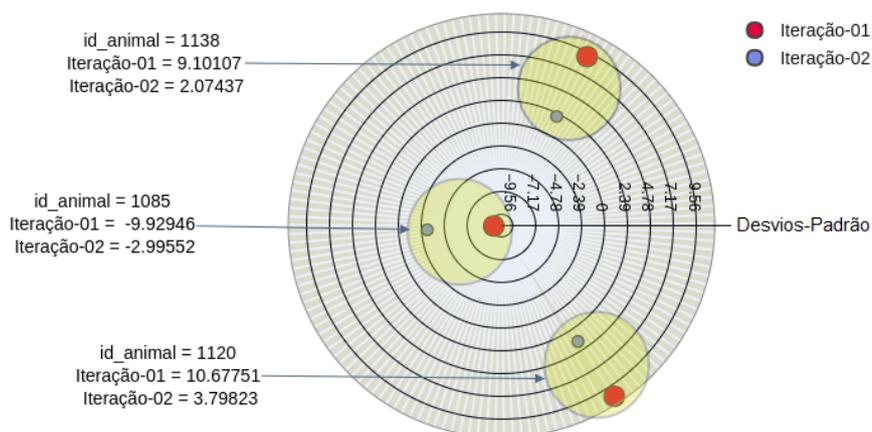
Animal_id	Desvio Padrão	Predição 1ª iteração	Predição 2ª iteração
1120	4,467	10.677511	3.79823
1138	3,808	9.101071	2.07437
1085	4,154	-9.929463	-2.99552

Fonte: Autor (2019)

Para representar a aproximação dos valores extremos à linha principal, a Figura 9 destaca somente os pontos que representam os *outliers*, realçados na Figura 7. A tendência de aproximação à linha principal pode ser percebida após a segunda iteração da MME. Os 3 pontos identificados como *outliers* foram distribuídos, de acordo com seus valores, considerando a primeira e segunda iterações, ou seja, a MME sem matriz  $D$  e, em seguida, com a matriz  $D$ . Os pontos referentes ao mesmo animal estão destacados, com a finalidade de demonstrar o deslocamento dos pontos na direção da linha com valor zero.

O processo pode ser repetido, aproximando ainda mais os resultados da linha principal, até que os valores, de iterações consecutivas, venham a convergir. Na Figura 8, apenas uma iteração foi suficiente para que os valores individuais do conjunto deixassem de ser considerados extremos. Entretanto, é factível ao pesquisador definir a quantidade de

Figura 9 – *Outliers* se aproximando da linha principal.

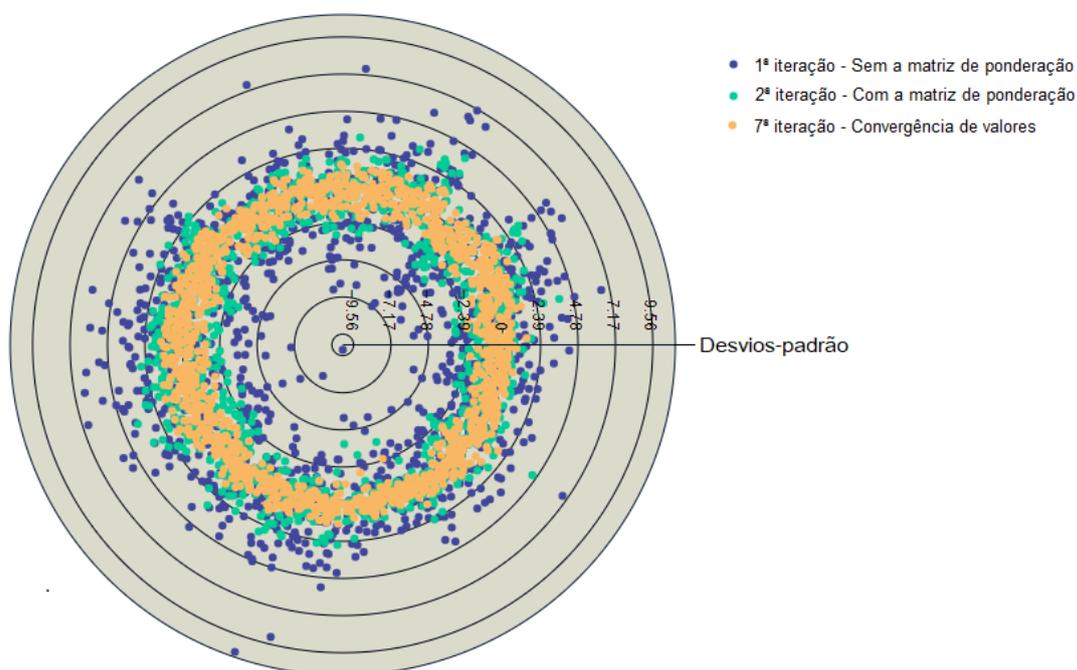


Fonte: Autor (2019)

desvios padrão aceitáveis, de acordo com o nível de confiança e necessidades da pesquisa. Da mesma forma, é possível definir que o algoritmo suspenda a execução ou prossiga, de maneira iterativa, mesmo após a eliminação dos *outliers*. Neste sentido, para demonstrar esta funcionalidade, foi definido que a etapa de cálculo do erro e recálculo da MME com a matriz  $D$  fosse repetida enquanto houvesse diferença entre os resultados dos grupos contemporâneos. Foi considerado um limite de duas casas decimais para que os resultados fossem considerados convergentes. A Figura 10 apresenta uma comparação entre os resultados iniciais, obtidos após a primeira execução, os resultados após a eliminação dos *outliers*, na primeira iteração do modelo e, os resultados após a convergência dos dados, ao final de sete iterações.

Fica demonstrado, através da Figura 10, que a introdução da matriz de ponderação no cálculo da MME foi capaz de mitigar as distorções geradas pelos valores extremos. Da mesma forma, a iteração do algoritmo, através do recálculo do erro residual e das variáveis de ponderação, apresentou um resultado mais satisfatório, no que diz respeito a concentração dos resultados próximos à linha principal. Para elucidar o processo de mitigação dos valores extremos foi selecionado um dos indivíduos, classificado como *outlier*, cujos valores são mostrados na Figura 11. O animal com id = 1138 foi considerado um *outlier* e o seu valor genético original, calculado pela MME, teve uma redução no fator de impacto. Desta forma, ao invés de computar 100% do valor genético, para a composição da média geral do seu grupo contemporâneo, o indivíduo em questão contribuiu com um percentual reduzido. Este percentual é proporcional ao desvio padrão encontrado e

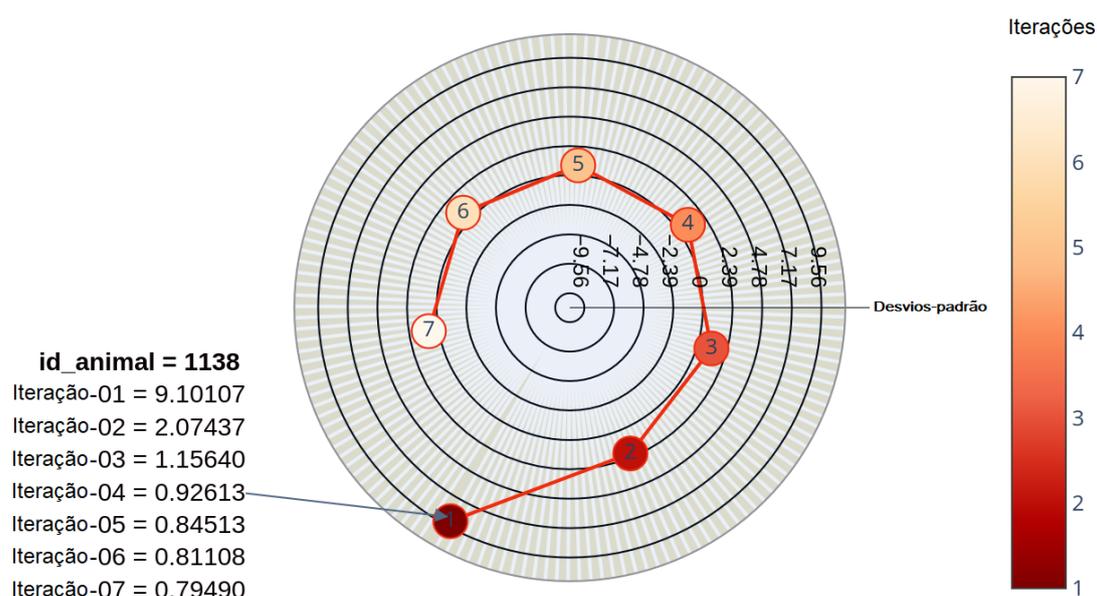
Figura 10 – Dispersão dos valores genéticos de todo o conjunto de indivíduos após o processamento.



Fonte: Autor (2019)

continua sendo reduzido a cada iteração, conforme mostra a Figura 11.

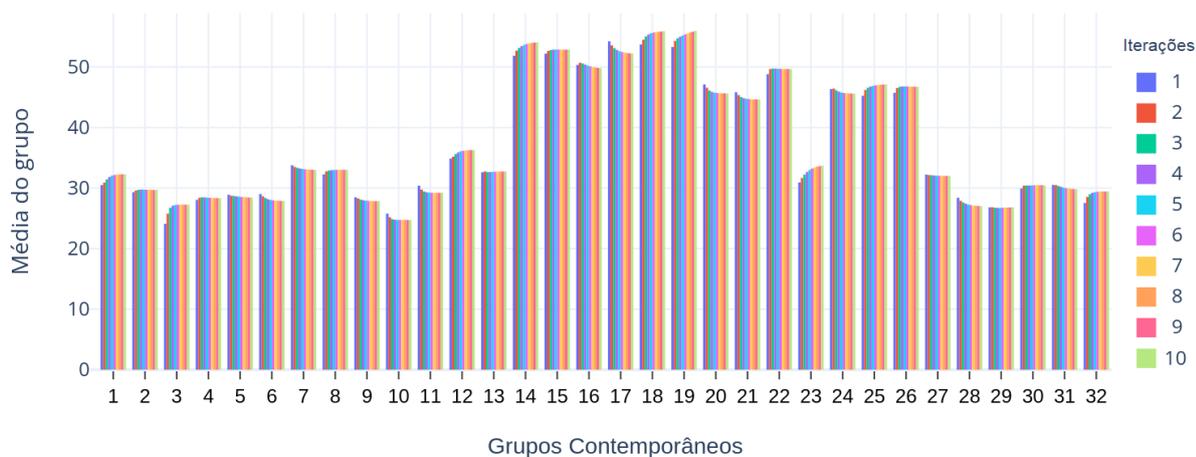
Figura 11 – Predição dos valores genéticos do animal id = 1183 a cada iteração.



Fonte: Autor (2019)

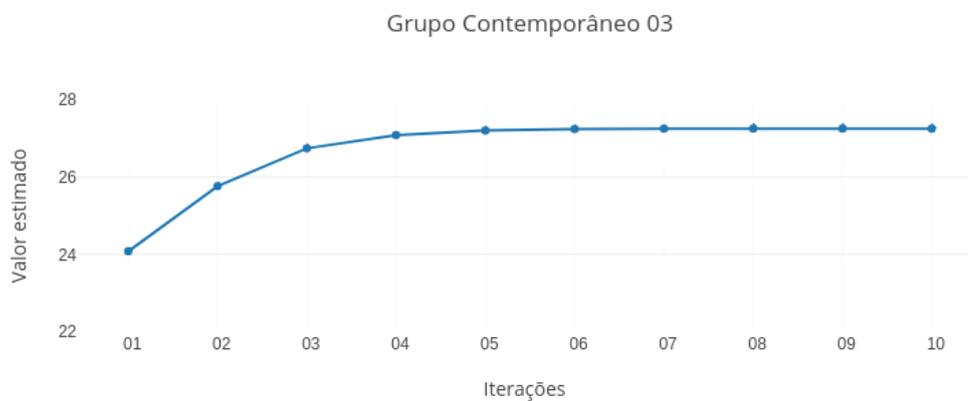
De maneira semelhante, os grupos contemporâneos também têm seus valores ajustados a cada iteração, devido à ponderação dos valores genéticos dos indivíduos que compõem o grupo. Os dados de AOL, analisados nesse estudo, estão distribuídos em 32 grupos contemporâneos, com quantidades disformes de indivíduos. A Figura 12 mostra, de maneira geral, a oscilação das médias de valores genéticos para todos os grupos contemporâneos em um intervalo de 10 iterações, onde cada barra vertical representa uma iteração. Em alguns grupos, a diferença entre a primeira e a última barra são quase imperceptíveis. Em outros, a diferença entre a primeira (I1) e última barra (I10) evidencia a oscilação no valor das médias do grupo. Esta oscilação será maior em grupos que possuem indivíduos com valores extremos, pois estes deixarão de fornecer o valor genético original, calculado pela MME e fornecerão o valor ponderado para a composição da média do grupo. As Figuras 13, 14 e 15 detalham as modificações ocorridas nos valores dos grupos contemporâneos 03, 12 e 29, respectivamente. A escolha destes grupos foi de maneira intencional, com a propósito de detalhar melhor a oscilação de valores dentro dos grupos, através de uma amostra contendo 3 situações distintas: O grupo contemporâneo com a maior alteração (gc03); outro, com proporções de alteração intermediárias (gc12); e um terceiro, com a menor alteração (gc29).

Figura 12 – Predição das médias dos valores genéticos separadas por grupos contemporâneos.



Fonte: Autor (2019)

Figura 13 – Predição das médias dos valores genéticos do grupo contemporâneo 03.



Fonte: Autor (2019)

Figura 14 – Predição das médias dos valores genéticos do grupo contemporâneo 12.



Fonte: Autor (2019)

Figura 15 – Predição das médias dos valores genéticos do grupo contemporâneo 29.



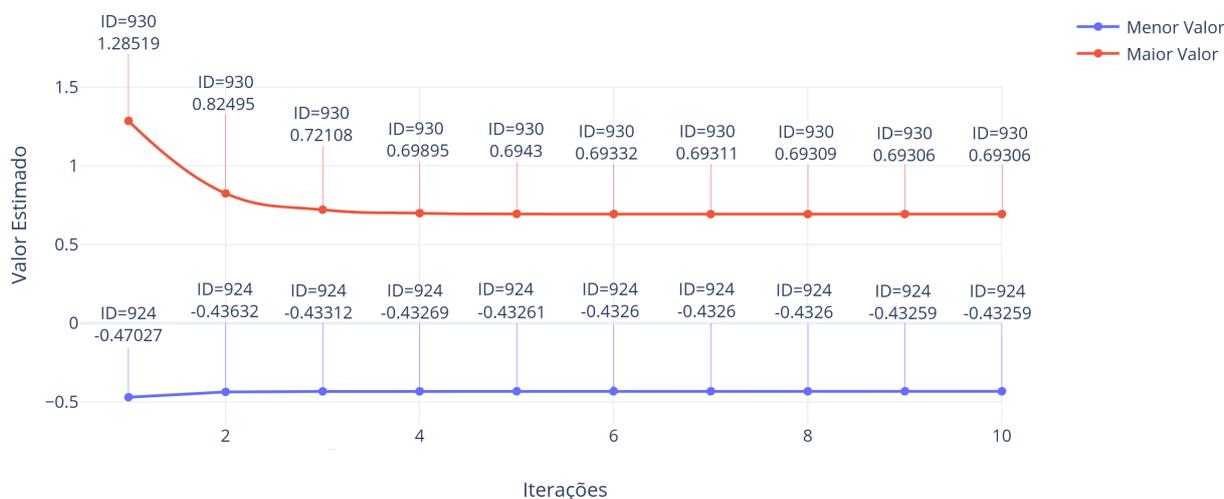
Fonte: Autor (2019)

## 4.2 Predição dos valores genéticos da espessura de gordura (EG)

A predição dos valores genéticos através da MME para a espessura de gordura (EG) considerou a mesma metodologia utilizada para AOL, apesar das bases de dados serem diferentes no quantitativo de animais. Para verificar se os dados seguiam uma distribuição normal foi aplicado o teste de normalidade de Shapiro-Wilk, onde foi verificado que o  $p$ -valor resultante foi  $2.2e-16$ , onde  $p < 0.05$  indica que os dados não possuem distribuição normal.

A influência dos *outliers*, na composição das médias dos grupos foi reduzida através do processo iterativo. Foram necessárias 10 iterações para que os resultados viessem a convergir em 5 casas decimais. A Figura 16 mostra um gráfico de amplitude, onde foram plotados o maior e o menor valores preditos para a EG, a cada iteração.

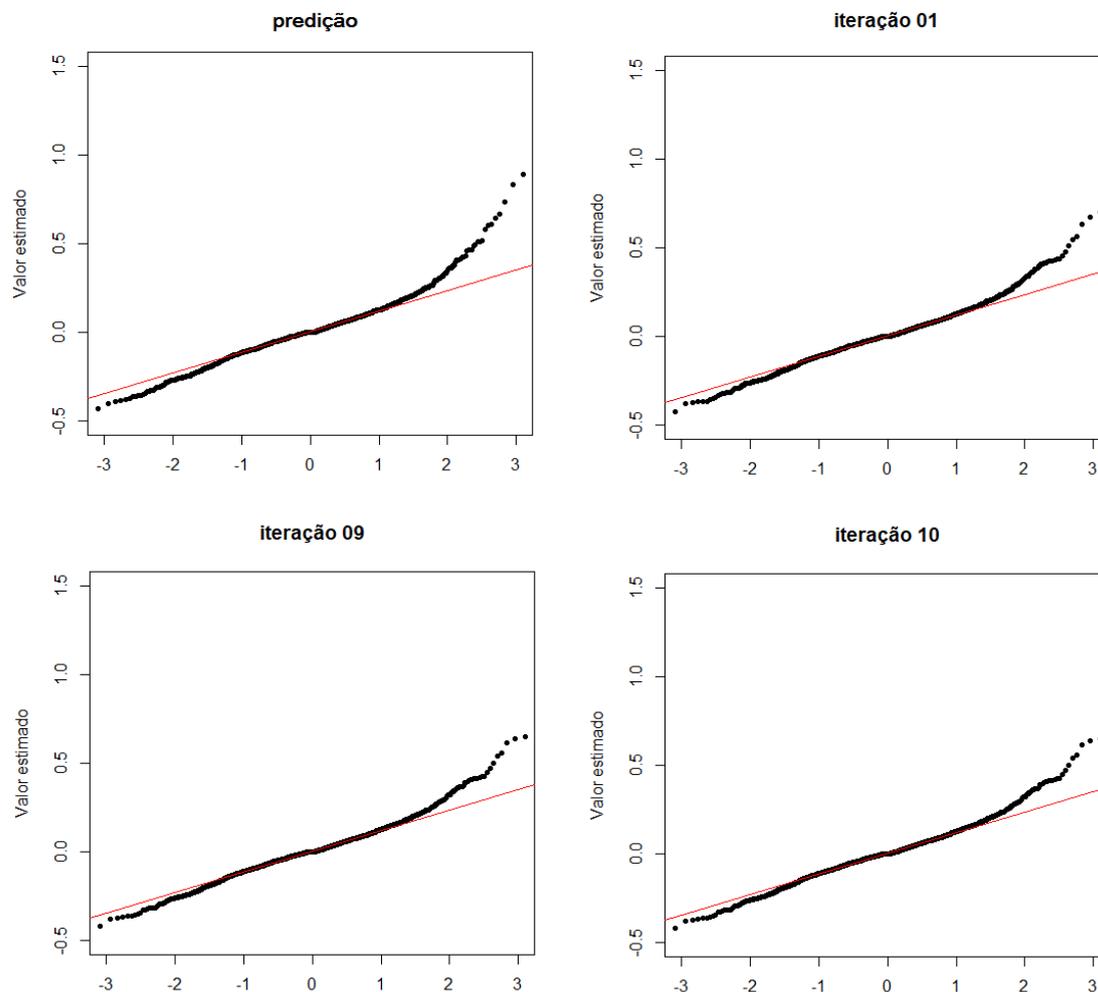
Figura 16 – Maior e menor valores individuais para a característica EG.



Fonte: Autor (2019)

Diferente da forma como foi apresentada a convergência dos valores de AOL, na Figura 10, a Figura 17 mostra um gráfico de dispersão para cada iteração do algoritmo, onde é possível observar o deslocamento dos pontos. Ao todo foram 10 iterações, representadas de forma resumida pelos dois primeiros e os dois últimos gráficos, onde é possível verificar que houve um deslocamento de pontos em direção ao eixo diagonal, traçado para representar a distribuição normal esperada. Os deslocamentos tornam-se menos significativos a cada iteração, até que deixam de existir.

Figura 17 – Dispersão dos valores genéticos de EG após cada iteração.



Fonte: Autor (2019)

A média dos valores estimados, após a correção dos *outliers*, mostrado pela Tabela 11 comprova que houve redução da amplitude, quando comparadas a primeira e a última estimações. Ou seja, diminuiu a diferença entre o menor e o maior valores estimados. No mesmo sentido, o desvio padrão que era de 0,15 diminuiu para 0,13 logo na primeira iteração, mantendo-se constante este valor até o fim da execução do algoritmo. As médias do BRBV foram apresentadas, na última linha, para validar o resultado da predição. A redução na amplitude é mostrada através da Figura 18, onde são dispostos 3 histogramas representando os valores estimados. No primeiro histograma a curva de normalidade, representada pela linha vermelha, possui caudas mais longas e vai de -0,47 até 1,29. No segundo histograma, que representa a primeira iteração, os valores vão de -0,42 até 0,70. No último histograma, quando os valores convergiram, a linha representa uma distribuição

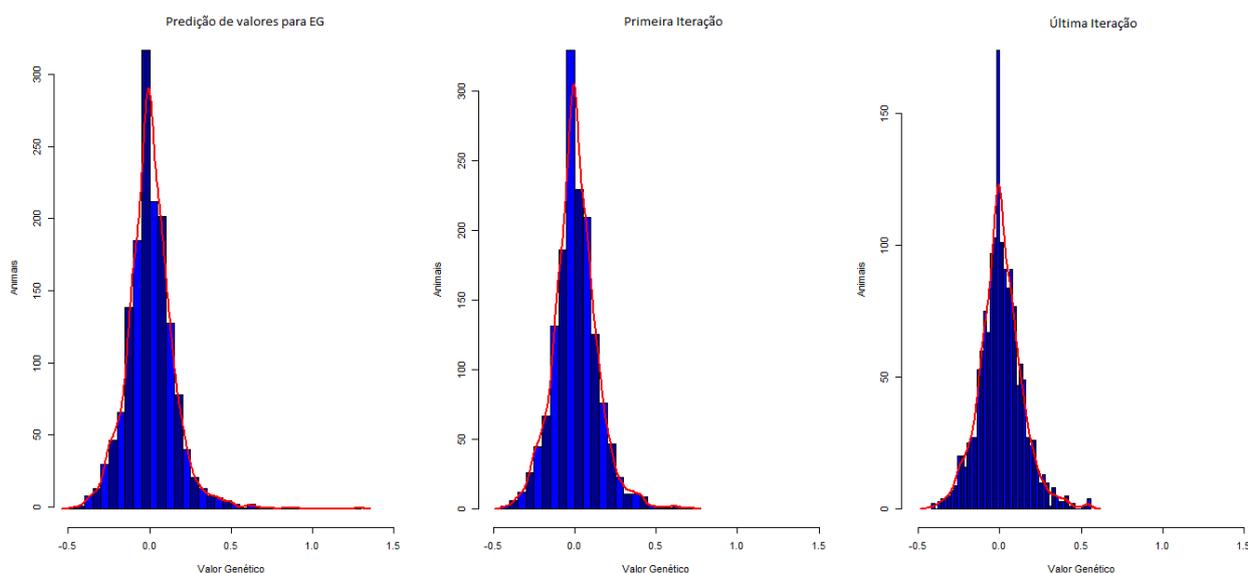
com a cauda mais curta, variando entre -0,42 e 0,55.

Tabela 11 – Comparação das médias dos valores estimados para EG após a última iteração.

	Quantidade	DP	Valor Mínimo	Valor Máximo
Predição	1552	0,15	-0.47	1.29
Primeira Iteração	1552	0,13	-0.42	0.70
Última Iteração	1552	0,13	-0.42	0.55
BRBV	1552	0,15	-0.47	1.29

Fonte: Autor (2019)

Figura 18 – Histogramas dos valores individuais preditos para EG.



Fonte: Autor (2019)

### 4.3 Validação dos resultados

Com o objetivo de assegurar a precisão dos resultados obtidos, através do cálculo das MME, foi considerado o BRBV como prova de conceito. O BRBV é um sistema, desenvolvido pela EMBRAPA, que implementa o BLUP e é capaz de estimar os valores genéticos de animais. Sua implantação consiste na instalação das dependências de *softwares* e bibliotecas, descritas na seção 2.7, que são pré-requisitos para a execução dos

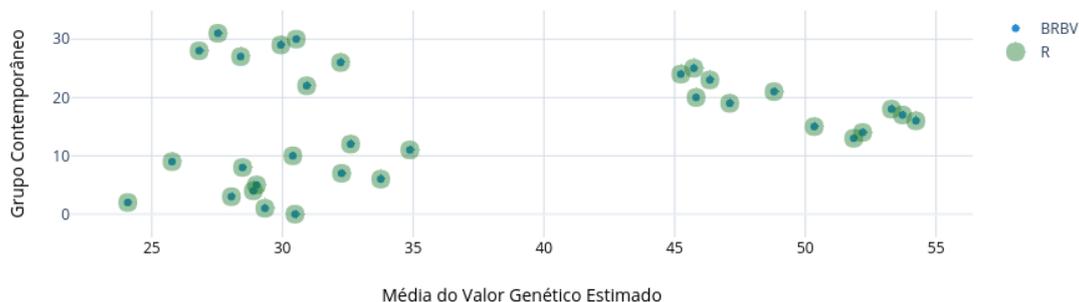
*scripts*. As demais características foram apresentadas na seção 2.7. Os mesmos dados, utilizados no modelo desenvolvido, foram importados e processados no sistema BRBV. Os resultados foram submetidos ao teste de correlação, onde verificou-se que eram equivalentes aos obtidos através da solução desenvolvida.

Para determinar a correlação entre os resultados obtidos pelas diferentes soluções, foi aplicado o teste de correlação linear de Pearson, onde é definido que os valores do coeficiente de correlação se situam entre os intervalos -1 e 1, assumindo que 0 significa a ausência de correlação entre as variáveis e  $\pm 1$  representa uma correlação perfeita, positiva ou negativa. De acordo com a classificação proposta por Pearson (1909), o coeficiente de correlação ( $R$ ) é interpretado da seguinte forma:

- $R = \pm 1$ : Quando a correlação é perfeita;
- $|0,91| \leq R < |1|$ : Correlação fortíssima;
- $|0,61| \leq R < |0,91|$ : Correlação forte;
- $|0,31| \leq R < |0,61|$ : Correlação média;
- $|0| < R < |0,31|$ : Correlação fraca;
- $R = 0$ : Quando a correlação é nula.

Diante disso, foram estabelecidas as matrizes resultantes da execução do BLUP como variáveis para o teste de correlação. Os valores individuais estimados, assim como as médias de cada grupo contemporâneo, foram considerados para o preenchimento das matrizes que foram submetidas ao teste de correlação. Estes valores são apresentados no apêndice B e, em ambos os casos, obteve-se o nível máximo de correlação. A distribuição das médias estimadas para os grupos contemporâneos existentes é mostrada na Figura 19, onde os pontos sobrepostos demonstram que os resultados obtidos utilizando o R são equivalentes aos obtidos com o BRBV. Para a produção das figuras, foram considerados os valores de AOL. A correlação perfeita entre os resultados é demonstrada através da Figura 20, onde são apresentados todos os valores individuais estimados. No eixo X estão dispostos os resultados obtidos utilizando a solução desenvolvida com o R, enquanto no eixo Y foram dispostos os valores estimados com o BRBV. O teste de correlação de Pearson para estas matrizes resultantes foi  $R = +1$ , ou seja, uma correlação perfeita entre os valores das matrizes.

Figura 19 – Média do valor genético da AOL estimado para cada grupo contemporâneo.



Fonte: Autor (2019)

Figura 20 – Correlação entre os valores genéticos individuais obtidos e a prova de conceito.



Fonte: Autor (2019)

#### 4.4 Acurácia

Os resultados apresentados nas seções 4.1 e 4.2, são predições do valor genético de cada animal para as características de interesse. A diferença entre este valor predito e o valor genético real é denominada acurácia, ou erro de predição. Desta forma, quanto maior for a acurácia, maior será a chance de que a DEP estimada corresponda ao valor genético real daquele indivíduo.

A acurácia é uma medida da correlação entre o valor estimado e o valor real, que expressa o quão confiável pode ser o resultado de uma predição. A medida de acurácia, recomendada pela *Beef Improvement Federation* (BIF) é a mais utilizada e se baseia na minimização da Variância do Erro de Predição (VEP) que é uma medida da magnitude dos erros na predição dos valores genéticos, obtida na diagonal da matriz intermediária do processo de solução das equações de modelos mistos e expressa através da Fórmula 7.

$$AC = 1 - \sqrt{\frac{VEP}{\sigma_a^2}} \quad (7)$$

onde:

- VEP = variância do erro da predição;
- $\sigma_a^2$  = variância genética aditiva da característica.

A tabela 12 mostra um resumo das médias da acurácia dos resultados. Na primeira coluna, são mostradas as médias após o cálculo da MME para a predição dos valores da EG. Nas colunas seguintes, são mostradas as médias das acurácias após o cálculo da primeira e da última iterações do modelo robusto. Os valores do BRBV são mostrados na última coluna, para fins de validação.

Tabela 12 – Resumo dos valores da acurácia das predições.

	Predição	Iteração01	Última Iteração	BRBV
Média da Acurácia	0,54	0,54	0,54	0,54
Desvio padrão	0,18	0,18	0,18	0,18
Mediana	0,62	0,62	0,62	0,62
Valor mínimo	0	0	0	0
Valor máximo	0,91	0,91	0,91	0,91

Fonte: Autor (2019)

Os valores da acurácia, para cada valor genético predito, foram comparados com os resultados do BRBV, através do teste de correlação de Pearson. O resultado, arredondado, foi  $R = +1$ . Indicando uma correlação perfeita entre os valores de acurácia calculados nos dois sistemas. Para validar os resultados obtidos com o modelo robusto, foi calculada a acurácia da predição em cada iteração e comparada com a acurácia obtida na primeira predição. A Tabela 13 apresenta os valores dos testes de correlação entre a predição inicial, primeira e última iterações do modelo robusto e a prova de conceito.

O cálculo das acurácias expressou, através de uma escala entre 0 e 1, a probabilidade do valor estimado permanecer inalterado em uma próxima avaliação. Os resultados foram equivalentes, em todas as iterações do modelo, pois o cálculo da acurácia depende da relação de consanguinidade entre os indivíduos e, portanto, mesmo modificando o va-

Tabela 13 – Correlação entre a acurácia da predição calculada, a acurácia da primeira e última iterações e a acurácia calculada no BRBV.

	Predição	Iteração01	Última Iteração	BRBV
Predição	1.0000000	0.9997142	0.9995728	0.9995306
Primeira Iteração	0.9997142	1.0000000	0.9999811	0.9992262
Última Iteração	0.9995728	0.9999811	1.0000000	0.9990846
BRBV	0.9995306	0.9992262	0.9990846	1.0000000

Fonte: Autor (2019)

lor do erro, para realizar o ajuste de *outliers*, a matriz de parentesco foi sempre a mesma.

Ao contrário dos cálculos de acurácia, as correlações entre os valores preditos diminui a cada iteração. Isso acontece devido ao ajuste dos *outliers*, que se aproximam da média a cada iteração. Na Tabela 14 é mostrada a correlação perfeita entre os valores preditos e suas prova de conceito. Também é mostrado que o valor da correlação diminui conforme aumentam as iterações.

Tabela 14 – Correlação entre a predição inicial do modelo, sua primeira e última iterações e a prova de conceito.

	Predicao	Iteração01	Última Iteração	BRBV
Predição	1.0000000	0.9886808	0.9799607	0.9999866
Primeira Iteração	0.9886808	1.0000000	0.9987087	0.9886882
Última Iteração	0.9799607	0.9987087	1.0000000	0.9799754
BRBV	0.9999866	0.9886882	0.9799754	1.0000000

Fonte: Autor (2019)

#### 4.5 Avaliação de desempenho

O desempenho computacional é uma variável que impacta na viabilização, ou não, da implantação de um algoritmo de predição robusta em um programa de melhoramento animal. O trabalho de Cardoso *et al.* (2009) aponta a demanda computacional como uma limitação para a implantação dos modelos baseados em MCMC. Neste mesmo sentido, Pereira (2004) avalia que a aplicação do método BLUP implica em um alto custo computacional devido à necessidade da inversão de uma matriz de ordem igual ao número de

animais x características.

Diante disso, buscou-se observar fatores capazes de contribuir com o desempenho do algoritmo desenvolvido. Dois fatores importantes foram observados na implementação deste algoritmo:

- A inversão de matrizes pelo método de substituição de Henderson (1976) evita o cálculo de inversão das matrizes e torna mais rápido o processamento do algoritmo;
- A comparação do tempo de execução, com o sistema BRBV, mostrou que, apesar da complexidade dos cálculos, o tempo de processamento foi adequado.

Para comparar tempo de execução foi acionado, no *shell* do linux, o comando que executa o BRBV, com os seus respectivos parâmetros, precedido do comando *time*, que fornece o tempo real de processamento destinado à tarefa. Em seguida, o mesmo foi feito em relação ao R. Foi utilizado o mesmo servidor, descrito na seção 3.2, para ambos os testes. Para o cálculo das MME, sem acurácia, foram contabilizadas 20 execuções consecutivas em cada sistema e os tempos de execução são mostrados na Figura 21, onde a média de tempo de execução para o algoritmo desenvolvido foi de 9,98 segundos, enquanto no BRBV o tempo médio foi de 27,75 segundos. Para o cálculo de acurácias, a diferença de tempo é mais significativa. Foi necessário, em média, 1 minuto para o algoritmo desenvolvido executar o cálculo das MME e, também, calcular as acurácias de todas as predições. No BRBV, o tempo médio para retornar o valor das predições e as respectivas acurácias foi de 8 horas e 3 minutos.

Figura 21 – Comparativo do tempo de processamento para o cálculo das MME.



Fonte: Autor (2019)

Neste capítulo foram apresentados os resultados obtidos através da implementação da metodologia proposta. Foi apresentada uma solução para as MME através da implementação do BLUP, em linguagem R, cuja validação dos resultados considerou, como prova de conceito, o sistema BRBV. Em seguida, foi abordada a problemática da predição de valores genéticos em conjuntos de dados com *outliers* e o aumento da precisão, nas estimações destes mesmos conjuntos, graças ao emprego do modelo robusto. No capítulo seguinte, são apresentadas as conclusões e demais considerações deste trabalho.

## 5 CONCLUSÕES

Foram implementadas as MME, através das quais foi possível identificar as relações existentes entre os valores extremos a precisão das estimações. Observou-se que conjuntos de dados com valores fora da curva normal tentem a enviesar o resultado das médias dos efeitos fixos, fazendo com que a DEP dos animais avaliados seja comparada com uma média não condizente com a realidade. A adequação do modelo, através da introdução de uma matriz de ponderação, foi capaz de corrigir o desvio de cada observação, mitigando a influência dos valores extremos e aproximando os resultados estimados dos resultados reais, melhorando assim a predição dos valores genéticos, em comparação ao verificado nos modelos mistos.

A solução desenvolvida consiste na implementação computacional do modelo robusto, onde nenhum dado é descartado, ou desprezado, devido ao valor discrepante. Entretanto, de forma proporcional, os dados tiveram o seu fator de impacto, na média do grupo, ajustados de acordo com o erro padrão calculado. Atribuindo uma escala de proporções, variando entre 0 e 1, os dados com valores estimados muito próximos dos valores reais foram multiplicados por uma variável cujo valor é proporcionalmente próximo de 1. De maneira semelhante, os valores estimados que estavam distantes da média geral, foram multiplicados por uma variável com valor proporcionalmente mais próximo de 0.

O processo iterativo, caracterizado pelo recálculo do erro padrão e variáveis de ponderação após a apuração dos resultados, foi capaz de aproximar com maior precisão os valores estimados dos valores reais, considerados para o cálculo da média dos grupos. Desta forma, após o cálculo das MME, os resultados apurados serviram como dados de entrada e puderam ser reprocessados até que os resultados viessem a convergir. A cada nova iteração os valores foram alterados, mas a proporção desta alteração foi cada vez menos significativa do ponto de vista do deslocamento dos resultados até a média.

Diante disso, baseado no fato de que a distorção dos resultados de avaliação, ocasionada pelos *outliers*, pode prejudicar o progresso genético das populações de bovinos de corte, sugerindo valores estimados sem a devida acurácia, é de fundamental importância que se considere esta interação nas avaliações genéticas que envolvem MME. Principalmente, quando se trata da análise de um conjunto massivo de dados, onde a diferença nos resultados pode ser pouco perceptível, dado o volume de informações e a proporção destas distorções, fazendo com que o material genético ou mesmo animais geneticamente inferiores tenham prioridade na seleção.

## REFERÊNCIAS

- AMARAL, T.; SERENO, J.; PELLEGRIN, A. **Fertilidade, funcionalidade e genética de touros zebuínos.**(2009). Corumbá: Embrapa Pantanal; Campo Grande, MS: Brasil, 2009.
- ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNES. Perfil da pecuária no brasil relatório anual 2018. **Associação Brasileira das Indústrias Exportadoras de Carnes (ABIEC)**, 2018.
- ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNES. Exportações brasileiras de carne bovina fecham 2018 com recorde histórico. **Associação Brasileira das Indústrias Exportadoras de Carnes (ABIEC)**, 2019.
- BARBOSA, J. J.; PEREIRA, T. M.; OLIVEIRA, F. L. P. de. Uma proposta para identificação de outliers multivariados. **Ciência e Natura**, v. 40, p. 40, 2018. Disponível em: <https://www.repositorio.ufop.br/handle/123456789/11454>. Acesso em: 25 ago. 2019.
- BOURDON, R. M.; BOURBON, R. M. **Understanding animal breeding**. Pearson: Prentice Hall Upper Saddle River, NJ, 2000.
- BRIDI, A. M. Qualidade da carne para o mercado internacional. **Departamento de Zootecnia Universidade Estadual de Londrina**, 2004. Disponível em: <http://www.uel.br/pessoal/ambridi/Carnesecarcacas.htm>. Acesso em: 25 ago. 2019.
- CARDOSO, F.; OLIVEIRA, M. de. Metodologia para estimação robusta com variâncias residuais heterogêneas em populações multirraciais. **Embrapa Pecuária Sul - Documentos (INFOTECA-E)**, 2015. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/136233/1/DT-118.pdf>. Acesso em: 25 ago. 2019.
- CARDOSO, F. F.; LOPA, T.; TEIXEIRA, B. **PampaPlus: Avaliação Genética Hereford e Braford**. Bagé: Embrapa Pecuária Sul, 2016.
- CARDOSO, F. F. *et al.* Modelos hierárquicos bayesianos para estimação robusta e análise de dados censurados em melhoramento animal. **Revista Brasileira de Zootecnia**, Sociedade Brasileira de Zootecnia, v. 38, n. sSuplemento Especial, 2009. Disponível em: <https://www.rbz.org.br/pt-br>. Acesso em: 25 ago. 2019.
- CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL. **O Futuro é Agro 2018-2023**. Brasil: Confederação da Agricultura e Pecuária do Brasil (CNA), 2018.
- COPT, S.; HERITIER, S. Robust alternatives to the f-test in mixed linear models based on mm-estimates. **Biometrics**, Wiley Online Library, v. 63, n. 4, p. 1045–1052, 2007. Disponível em: <https://www.jstor.org/publisher/ibs>. Acesso em: 25 ago. 2019.
- COPT, S.; VICTORIA-FESER, M.-P. High-breakdown inference for mixed linear models. **Journal of the American Statistical Association**, Taylor & Francis, v. 101, n. 473, p. 292–300, 2006. Disponível em: <https://amstat.tandfonline.com>. Acesso em: 25 ago. 2019.

COSTA, D. R. *et al.* **Estimação robusta em modelos de variáveis latentes para dados censurados**. Tese (Doutorado) — Instituto de Matemática, Estatística e Computação Científica - UNICAMP, 2013.

CRUZ, C. **Programa Genes: diversidade genética**. Brasil: UFV, 2008.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Embrapa em números**. Brasília, Brasil: EMBRAPA, 2017.

EUCLIDES FILHO, K. **Produção de bovinos de corte e o trinômio genótipo-ambiente-mercado**. Campo Grande, MS.: Embrapa Gado de Corte-Documents (INFOTECA-E), 2000.

FARIA, C. U. d. *et al.* Inferência bayesiana e sua aplicação na avaliação genética de bovinos da raça nelore. **Ciência Animal Brasileira**, Brasil, 2007. Disponível em: <https://www.revistas.ufg.br/vet/article/view/1161>. Acesso em: 25 ago. 2019.

FERRAZ, J. B. S.; ELER, J. P. Parceria público x privada no desenvolvimento de pesquisa em melhoramento genético animal. **Revista Brasileira de Zootecnia**, Sociedade Brasileira de Zootecnia, v. 39, n. 4, p. 216–222, 2010. Disponível em: <https://www.rbz.org.br/ptbr>. Acesso em: 25 ago. 2019.

GIANOLA, D. *et al.* Prediction of complex traits: robust alternatives to best linear unbiased prediction. **Frontiers in genetics**, v. 9, p. 195, 2018. Disponível em: <https://www.frontiersin.org/journals/genetics>. Acesso em: 25 ago. 2019.

GIANOLA, D.; IM, S.; MACEDO, F. A framework for prediction of breeding value. In: **Advances in statistical methods for genetic improvement of livestock**. Springer: Nature, 1990. p. 210–238.

GIL, A. C. **Como elaborar projetos de pesquisa**. Brasil: Atlas, 2002.

GILL, P. S. A robust mixed linear model analysis for longitudinal data. **Statistics in medicine**, Wiley Online Library, v. 19, n. 7, p. 975–987, 2000. Disponível em: <https://onlinelibrary.wiley.com/toc/10970258/2000/19/7>. Acesso em: 25 ago. 2019.

HENDERSON, C. Best linear unbiased prediction of breeding values not in the model for records. **Journal of Dairy Science**, Elsevier, v. 60, n. 5, p. 783–787, 1977. Disponível em: [https://www.journalofdairyscience.org/article/S0022-0302\(77\)83935-0/pdf](https://www.journalofdairyscience.org/article/S0022-0302(77)83935-0/pdf). Acesso em: 25 ago. 2019.

HENDERSON, C. R. Sire evaluation and genetic trends. **Journal of Animal Science**, Oxford University Press, v. 1973, n. Symposium, p. 10–41, 1973. Disponível em: <https://doi.org/10.1093/ansci/1973.Symposium.10>. Acesso em: 25 ago. 2019.

HENDERSON, C. R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. **Biometrics**, JSTOR, p. 69–83, 1976. Disponível em: <https://www.jstor.org/stable/2529339>. Acesso em: 25 ago. 2019.

JAMROZIK, J.; FATEHI, J.; SCHAEFFER, L. Application of robust procedures for estimation of breeding values in multiple-trait random regression test-day model. **Journal of animal breeding and genetics**, Wiley Online Library, v. 124, n. 1, p. 3–11, 2007. Disponível em: <https://onlinelibrary.wiley.com/toc/14390388/2007/124/1>. Acesso em: 25 ago. 2019.

JIANG, J. **Linear and generalized linear mixed models and their applications**. USA: Springer Science & Business Media, 2007.

LÔBO, R. N. B.; LÔBO, A. Melhoria genética como ferramenta para o crescimento e o desenvolvimento da ovinocultura de corte. In: **Embrapa Caprinos e Ovinos-Artigo em periódico indexado (ALICE)**. v. 31, n. 2, p. 247-253. Belo Horizonte: Revista Brasileira de Reprodução Animal, 2007. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/33204/1/API-Melhoramento-genetico-como-ferramenta.pdf>. Acesso em: 25 ago. 2017.

MAIA, M. C. C. *et al.* Seleção de clones experimentais de cupuaçu para características agroindustriais via modelos mistos. **Revista Brasileira de Agroambiente**, v. 5, n. 1, 2011. Disponível em: <https://revista.ufr.br/agroambiente/article/download/388/431>. Acesso em: 25 ago. 2019.

MONTALDO, H. H.; MEZA-HERRERA, C. A. Use of molecular markers and major genes in the genetic improvement of livestock. **Electronic Journal of Biotechnology**, Pontificia Universidad Católica de Valparaíso, v. 1, n. 2, p. 15–16, 1998. Disponível em: <http://www.bioline.org.br/pdf?ej98011>. Acesso em: 25 ago. 2019.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. 3. ed. USA: Cabi, 2014.

NOGUEIRA, D. A. *et al.* Análises clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Ciência e Agrotecnologia, Lavras**, v. 27, p. 1614–1624, 2003. Disponível em: <http://www.editora.ufla.br>. Acesso em: 25 ago. 2019.

PEARSON, K. Determination of the coefficient of correlation. **Science**, JSTOR, v. 30, n. 757, p. 23–25, 1909.

PEREIRA, C. d. F. *et al.* **A Importância da qualidade da informação na predição de valores genéticos para características de crescimento em bovinos da raça nelore**. Dissertação (Mestrado) — Universidade Federal de Uberlândia, 2014. Disponível em: <https://doi.org/10.1590/1678-4162-8942>. Acesso em: 25 ago. 2019.

PEREIRA, J. C. C. **Melhoramento genético aplicado à produção animal**. Brasil: FEP MVZ, 2004.

PÉREZ, P.; CAMPOS, G. de L. Genome-wide regression and prediction with the bglr statistical package. **Genetics**, Genetics Soc America, v. 198, n. 2, p. 483–495, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4196607>. Acesso em: 25 ago. 2019.

PIRES, A. M.; BRANCO, J. A. **Introdução aos métodos estatísticos robustos**. Brasil: Congresso Anual da Sociedade Portuguesa de Estatística, 2007.

POLAQUINI, L. E. M.; SOUZA, J. G. d.; GEBARA, J. J. Transformações técnico-produtivas e comerciais na pecuária de corte brasileira a partir da década de 90. **Revista Brasileira de Zootecnia**, SciELO Brasil, p. 321–327, 2006. Disponível em: <https://doi.org/10.1590/S1516-35982006000100040>. Acesso em: 25 ago. 2019.

- POLLACK, E. Expected progeny differences within breed comparisons. **Journal of Animal Science**, p. 17–23, 1992. Disponível em: <http://www.journalofanimalscience.org/content/83/10/2461>. Acesso em: 25 ago. 2019.
- RESENDE, M. *et al.* Melhor predição linear não viciada (blup) de valores genéticos no melhoramento de pinus. **Embrapa Florestas-Artigo em periódico indexado (ALICE)**, Boletim de Pesquisa Florestal, Colombo, n. 32/33, p. 3-22., 1996. Disponível em: <https://www.alice.cnptia.embrapa.br/handle/doc/282155>. Acesso em: 25 ago. 2019.
- RESENDE, M. d. *et al.* **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial**. Viçosa Brasil: Ed. UFV, 2012.
- RICHARDSON, A. M. Bounded influence estimation in the mixed linear model. **Journal of the American Statistical Association**, Taylor & Francis, v. 92, n. 437, p. 154–161, 1997. Disponível em: <https://www.jstor.org/stable/2291459?seq=1>. Acesso em: 25 ago. 2019.
- RICHARDSON, A. M.; WELSH, A. H. Robust restricted maximum likelihood in mixed linear models. **Biometrics**, JSTOR, p. 1429–1439, 1995. Disponível em: <https://www.jstor.org/stable/2533273>. Acesso em: 25 ago. 2019.
- ROSA, A. d. N. **Melhoramento genético aplicado em gado de corte**. Campo Grande, MS, Brasil: Empresa Brasileira de Pesquisa Agropecuária, 2018.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, JSTOR, v. 52, n. 3/4, p. 591–611, 1965. Disponível em: <https://www.jstor.org/stable/2333709>. Acesso em: 25 ago. 2019.
- SIEVERT, C. *et al.* plotly: Create interactive web graphics via plotly. js. **R package version**, v. 4, n. 1, p. 110, 2017.
- SIROL, M. L. F. G. *et al.* Effect of outliers in contemporary groups on breeding value prediction. **Revista Brasileira de Zootecnia**, SciELO Brasil, v. 36, n. 4, p. 807–812, 2007. Disponível em: <https://www.rbz.org.br/ptbr>. Acesso em: 25 ago. 2019.
- STRANDÉN, I.; GIANOLA, D. Mixed effects linear models with t-distributions for quantitative genetic analysis: a bayesian approach. **Genetics Selection Evolution**, BioMed Central, v. 31, n. 1, p. 25, 1999. Disponível em: <https://doi.org/10.1186/1297-9686-31-1-25>. Acesso em: 25 ago. 2019.
- TEAM, R. C. *et al.* R: A language and environment for statistical computing. Vienna, Austria, 2013.
- VAZQUEZ, A. *et al.* an r package for fitting generalized linear mixed models in animal breeding. **Journal of animal science**, Oxford University Press, v. 88, n. 2, p. 497–504, 2010. Disponível em: <https://doi.org/10.2527/jas.2009-1952>. Acesso em: 25 ago. 2019.
- VICENSOTTI, J. M.; MONTEBELLO, A. E. S.; MARJOTTA-MAISTRO, M. C. Competitividade brasileira no comércio exterior da carne bovina. **Revista IPecege**, v. 5, n. 1, p. 7–18, 2019. Disponível em: <https://doi.org/10.22167/r.ipecege.2019.5.7>. Acesso em: 25 ago. 2019.

WAZLAWICK, R. **Metodologia de pesquisa para ciência da computação**. Brasil: Editora Campus, 2017.

WELSH, A.; RICHARDSON, A. 13 approaches to the robust estimation of mixed models. **Handbook of statistics**, Science Direct, v. 15, p. 343–384, 1997. Disponível em: [https://doi.org/10.1016/S0169-7161\(97\)15015-5](https://doi.org/10.1016/S0169-7161(97)15015-5). Acesso em: Ago/2019.

WICKHAM, H. ggplot2. **Wiley Online Library**, v. 3, n. 2, p. 180–185, 2011. Disponível em: <https://doi.org/10.1002/wics.147>. Acesso em: 25 ago. 2019.

WIMMER, V. *et al.* synbreed: a framework for the analysis of genomic prediction data using r. **Bioinformatics**, Oxford University Press, v. 28, n. 15, p. 2086–2087, 2012. Disponível em: <https://doi.org/10.1093/bioinformatics/bts335>. Acesso em: 25 ago. 2019.

YAMANAKA, S. E. *et al.* **Modelos não Lineares Mistos em Estudos de Crescimento de Frango de Corte**. Dissertação (Mestrado) — Universidade Federal de São Carlos, 2018. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/10380>. Acesso em: 25 ago. 2019.

YOKOO, M. J. I. **Estimativas de efeitos genéticos e ambientais para características de carcaça mediadas pelo ultra-som em bovinos da raça nelore**. Dissertação (Mestrado) — Universidade Estadual Paulista (UNESP), 2005. Disponível em: <http://hdl.handle.net/11449/92595>. Acesso em: 25 ago. 2019.

YOKOO, M. J.-I. *et al.* O uso da estatística bayesiana no melhoramento genético animal: uma breve explicação. **Embrapa Pecuária Sul - Artigo em periódico indexado (ALICE)**, Scientia Agraria Paranaensis, Marechal Cândido Rondon, v. 12, n. 4, p. 247 , 2013. Disponível em: <http://www.alice.cnptia.embrapa.br/alice/handle/doc/974791>. Acesso em: 25 ago. 2019.

## APÊNDICE A — CÁLCULO DO BLUP PARA A PREDIÇÃO DE GANHO DE PESO À DESMAMA E EFEITOS FIXOS DO SEXO

O exemplo a seguir foi discutido no capítulo 2.3 e consiste em um modelo para avaliação de animais, baseado em Mrode (2014), utilizando um conjunto de dados amostrais, representados pela Tabela 1, onde são apresentados os valores de ganho de peso, pós-desmame, *Weaning Weight Gain* (WWG) de um grupo de animais. O objetivo desta avaliação é demonstrar a utilização do método BLUP, para estimar os efeitos do sexo e prever valores genéticos para todos os animais e a equação matemática que descreve este modelo animal é a seguinte:

$$\gamma = X\beta + Z\alpha + \varepsilon$$

Onde:

- $\gamma$ : ( $n \times 1$ ) vetor de observações;
- $\beta$ : ( $p \times 1$ ) vetor dos efeitos fixos;
- $\alpha$ : ( $q \times 1$ ) vetor dos efeitos aleatórios;
- $\varepsilon$ : ( $n \times 1$ ) vetor de erros aleatórios;
- $X$ : matriz de incidência ( $n \times p$ ) que relaciona registros a efeitos fixos;
- $Z$ : matriz de incidência ( $n \times q$ ) que relaciona registros a efeitos aleatórios;
- $n$ : número de registros;
- $p$ : número de níveis para efeitos fixos;
- $q$ : número de níveis para efeitos aleatórios.

Iniciou-se a construção do modelo pela definição das matrizes de incidência, neste caso,  $X$  e  $Z$ . A matriz  $X$ , neste modelo, de acordo com a notação apresentada, relaciona os registros com efeitos fixos (sexo). Para o conjunto de dados de exemplo, a matriz de incidência  $X$  já representada em sua forma transposta  $X'$  é a matriz de incidência do sexo sobre o animal, onde a primeira linha indica que a primeira, quarta e quinta observações são de animais do sexo masculino e a segunda linha indica que a segunda e terceira

observações são do sexo feminino.

$$X' = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

A matriz  $Z$  incide sobre todos os animais, independente de possuir ou não possuir registro de rendimento. Pode-se observar, através da distribuição dos valores na matriz, que as três primeiras colunas possuem apenas valores zerados, pois os animais 1, 2 e 3 não possuíam registro de rendimento.

$$Z = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Finalmente, o valor de  $\gamma$ , que corresponde ao efeito que se pretende analisar (WWG) permanece inalterado e, para efeito de cálculo, pode ser representado pelo vetor a seguir:

$$\gamma = \begin{pmatrix} 4.5 & 2.9 & 3.9 & 3.5 & 5 \end{pmatrix}$$

Neste ponto, foram obtidas as matrizes de entrada, a matriz transposta do efeito do sexo ( $X'$ ) e a matriz de animais com dados ( $Z$ ). As outras matrizes do modelo,  $X'Z$ ,  $Z'X$ ,  $X'\gamma$  e  $Z'\gamma$  foram obtidas através da multiplicação das primeiras matrizes e também do vetor  $\gamma$ . O procedimento consiste na multiplicação de coluna por linha. O quadro 1 demonstra o processo utilizado para calcular  $X'Z$ .

Quadro1: Detalhamento do cálculo para multiplicação da matriz  $XZ$

$1*0+0*0+0*0+1*0+1*0$	$1*0+0*0+0*0+1*0+1*0$	$1*0+0*0+0*0+1*0+1*0$	$1*1+0*0+0*0+1*0+1*0$
$1*0+0*1+0*0+1*0+1*0$	$1*0+0*0+0*1+1*0+1*0$	$1*0+0*0+0*0+1*1+1*0$	$1*0+0*0+0*0+1*0+1*1$
$0*0+1*0+1*0+0*0+0*0$	$0*0+1*0+1*0+0*0+0*0$	$0*0+1*0+1*0+0*0+0*0$	$0*1+1*0+1*0+0*0+0*0$
$0*0+1*1+1*0+0*0+0*0$	$0*0+1*0+1*1+0*0+0*0$	$0*0+1*0+1*0+0*1+0*0$	$0*0+1*0+1*0+0*0+0*1$

$$\begin{array}{ccc}
 X' & & Z \\
 \left( \begin{array}{cccc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array} \right) & \times & \left( \begin{array}{cccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) = \left( \begin{array}{cccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{array} \right) \\
 & & X'Z
 \end{array}$$

A transposição de matrizes é o procedimento através do qual a representação de uma matriz é invertida entre suas dimensões, sem alteração nos valores. A matriz  $Z'X$ , corresponde à matriz  $X'Z$  transposta. Desta forma, ao multiplicar a matriz  $X'$  pela matriz  $Z$ , obteve-se:

$$X'Z = \left( \begin{array}{cccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{array} \right) \quad Z'X = \left( \begin{array}{c} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{array} \right)$$

$X'\gamma$ , assim como  $Z'\gamma$ , é obtido através da soma dos valores de  $\gamma$  para cada valor válido em  $X'$  ou  $Z'$ . Desta forma, obteve-se:

$$X'\gamma = \left( \begin{array}{c} 13 \\ 6.8 \end{array} \right) \quad Z'\gamma = \left( \begin{array}{c} 0 \\ 0 \\ 0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5 \end{array} \right)$$

Para a estimação dos valores dos efeitos, é utilizado um sistema de equações lineares definido pela matriz de equações exposta a seguir. O resultado de cada multiplicação é concatenado aos demais, na formação de uma única matriz, onde  $XX$  pode ser verificado como o primeiro conjunto de 2 x 2 elementos, na parte superior esquerda. Ao seu lado

direito, encontra-se o conjunto de elementos  $XZ$  dispostos em 2 linhas e 8 colunas.  $ZX$  é representado pelas duas primeiras colunas e as 8 últimas linhas e, finalmente  $ZZ$  que é o conjunto mais representativo, está disposto na parte inferior direita, uma malha numérica de  $8 \times 8$ , iniciando na terceira linha e terceira coluna, onde pode-se observar a quase totalidade de elementos zerados, exceto pela diagonal principal, composta pelo valor 1, com exceção dos três primeiros, onde não existem valores registrados para os animais 1, 2 e 3 no modelo.

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix} \times \begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} X'\gamma \\ Z'\gamma \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{pmatrix} = \begin{pmatrix} 13 \\ 6.8 \\ 0 \\ 0 \\ 0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5 \end{pmatrix}$$

A matriz de parentesco terá dimensões  $n \times n$ , em que  $n$  corresponde ao número de indivíduos e, onde cada elemento corresponde ao grau de parentesco entre estes indivíduos. Seu preenchimento é dado de forma que os elementos na diagonal principal representem o grau de semelhança de indivíduos com eles mesmos, devendo, seus valores, serem iguais a soma de 1 mais o coeficiente de endogamia. Os elementos fora da diagonal principal representam os numeradores do coeficiente de parentesco entre os indivíduos.

Foi considerado o grau de parentesco determinado por Mrode (2014), e descrito na seção 2.4 que atribui pesos diferenciados para relação de consanguinidade, considerando a probabilidade percentual de serem encontrados alelos em comum. Foram atribuídos: 0.5 para o relacionamento entre pais e filhos 0.5 para o relacionamento entre irmãos 0.25 para o relacionamento entre meio-irmãos Estes valores são estimativas médias e podem variar. Para parentescos em gerações seguintes, considera-se a atribuição metade da pontuação atribuída.

A matriz de parentesco para este exemplo foi calculada, utilizando o *software* R, assim como a inversão da matriz, conforme mostrado na seção 2.4 seguindo os preceitos de (HENDERSON, 1976). A matriz inversa, introduzida no modelo conforme Mrode (2014) considera os oito animais do conjunto e suas relações de parentesco calculadas, dispostas em uma matriz  $A^{-1}$  de 8 x 8 com os seguintes valores:

$$A^{-1} = \begin{pmatrix} 1.833 & 0.5 & 0 & -0.667 & 0 & -1 & 0 & 0 \\ 0.5 & 2 & 0.5 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0.5 & 2 & 0 & -1 & 0.5 & 0 & -1 \\ -0.667 & 0 & 0 & 1.833 & 0.5 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0.5 & 2.5 & 0 & -1 & 0 \\ -1 & -1 & 0.5 & 0 & 0 & 2.5 & 0 & -1 \\ 0 & 0 & 0 & -1 & -1 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & 2 \end{pmatrix}$$

O cálculo da inversa de  $A$ , para a obtenção da matriz de parentesco foi incorporado ao modelo e é representado pela equação seguinte. A concatenação dos valores para a composição da matriz foi feita do mesmo modo, substituindo-se apenas, o conteúdo de  $Z'Z$  por  $Z'Z + A^{-1}$

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1} \end{pmatrix} \times \begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} X'\gamma \\ Z'\gamma \end{pmatrix}$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 3.666 & 1 & 0 & -1.334 & 0 & -2 & 0 & 0 \\ 0 & 0 & 1 & 4 & 1 & 0 & -2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 4 & 0 & -2 & 1 & 0 & -2 \\ 1 & 0 & -1.334 & 0 & 0 & 4.666 & 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & -2 & -2 & 1 & 6 & 0 & -2 & 0 \\ 0 & 1 & -2 & -2 & 1 & 0 & 0 & 6 & 0 & -2 \\ 1 & 0 & 0 & 0 & 0 & -2 & -2 & 0 & 5 & 0 \\ 1 & 0 & 0 & 0 & -2 & 0 & 0 & -2 & 0 & 5 \end{pmatrix}^{-1} \times \begin{pmatrix} 13 \\ 6.8 \\ 0 \\ 0 \\ 0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{pmatrix} = \begin{pmatrix} 4.3584 \\ 3.4044 \\ 0.0984 \\ -0.0187 \\ -0.0410 \\ -0.0086 \\ -0.1857 \\ 0.1768 \\ -0.2494 \\ 0.1826 \end{pmatrix}$$

Depois de calcular o modelo misto, o R gera um vetor de resultados de onde podem ser extraídos os efeitos fixos do sexo e os valores genéticos de cada indivíduo. O vetor apresentado é o resultado da equação, onde as duas primeiras linhas são os efeitos fixos de sexo, ou seja, a média dos valores genéticos para cada um dos grupos. Neste caso, machos na primeira linha e fêmeas na segunda linha. As demais linhas, correspondem aos valores genéticos de cada indivíduo. O resultado indica que os bezerros do grupo  $\beta_1$  têm uma propensão a ganhar 4,35kg de peso diário após o desmame. Da mesma forma, os bezerros do grupo  $\beta_2$  tiveram seu ganho diário de peso após o desmame estimado em 3,40 kg. Além disso, é possível inferir o valor genético individual para cada animal. As últimas linhas do vetor apresentam este resultado, considerando a diferença de ganho de peso ao desmame entre cada animal comparado com a média do seu grupo.

## APÊNDICE B — VALORES GENÉTICOS ESTIMADOS PARA A CARACTERÍSTICA DE AOL, SUBMETIDOS À PROVA DE CONCEITO

Valores genéticos estimados para cada um dos grupos contemporâneos:

[GC01] 30.47747  
 [GC02] 29.32572  
 [GC03] 24.07999  
 [GC04] 28.04433  
 [GC05] 28.88385  
 [GC06] 29.00497  
 [GC07] 33.75782  
 [GC08] 32.25876  
 [GC09] 28.47141  
 [GC10] 25.77346  
 [GC11] 30.39308  
 [GC12] 34.87035  
 [GC13] 32.60485  
 [GC14] 51.85041  
 [GC15] 52.19117  
 [GC16] 50.33938  
 [GC17] 54.22772  
 [GC18] 53.71666  
 [GC19] 53.29405  
 [GC20] 47.10843  
 [GC21] 45.81959  
 [GC22] 48.80273  
 [GC23] 30.92198  
 [GC24] 46.34589  
 [GC25] 45.23607  
 [GC26] 45.73131  
 [GC27] 32.22156  
 [GC28] 28.39832  
 [GC29] 26.81004  
 [GC30] 29.93636  
 [GC31] 30.52414  
 [GC32] 27.52662

Valores genéticos individuais estimados para o conjunto de animais:

[1] -5.55230 1.70730 0.00000 0.77019 1.28271 1.28271 -0.21825 2.56542 0.00000 0.00000 3.41460 0.00000 0.00000 2.67302 0.00000 -1.09730  
 [17] 1.69636 0.00000 -0.38215 3.14521 0.00000 -0.83871 -0.21086 1.31176 0.69789 1.87907 -5.02218 0.68266 0.00000 -2.59141 0.68995 0.72875  
 [33] 0.00000 -2.47419 1.99837 -2.28331 0.49147 1.60797 0.86446 0.52922 -1.05613 0.76935 0.90528 1.54038 0.00000 0.33724 0.41603 -0.93412  
 [49] 0.00000 1.91545 0.47162 0.49768 -4.52201 -1.48837 -0.02966 -1.24275 -1.04842 -2.97465 -1.26944 -0.55365 0.00000 -0.31892 -1.79974 0.34521  
 [65] 2.25949 0.70851 -0.59236 0.85116 -1.23288 0.00000 -5.33017 2.11557 1.65716 1.14457 0.00000 -2.35866 0.00000 -2.79203 -2.34428 -2.83127  
 [81] -1.92640 -0.80074 0.12971 2.14808 -0.81373 0.00000 0.00000 0.21814 1.84590 0.26625 0.82858 -0.30959 -0.05501 0.81217 0.03973 -2.30804  
 [97] -3.22866 1.28011 2.17188 0.00000 0.00000 0.00000 1.57202 3.91053 0.00000 1.18781 -2.76303 -1.32059 1.58596 2.42685 0.00000 4.31294  
 [113] 0.00000 0.00000 2.58919 3.41460 0.63741 0.00000 -1.52950 -1.22919 0.00000 1.08487 2.70052 1.42992 1.51877 -0.46521 -2.67364 -0.84440  
 [129] -1.72275 0.00000 0.00000 0.34569 0.00000 1.72544 0.36364 0.47311 -0.11705 1.19804 -5.03362 0.04652 -2.17084 0.00000 0.24573 0.49463  
 [145] -0.35393 -2.15970 -0.03927 -2.65099 3.34261 0.00000 -0.59609 1.40052 1.24404 0.17095 2.34001 -1.00914 -2.94571 0.00000 3.37242 1.73693  
 [161] -0.38830 -2.57676 -2.76014 -0.02105 3.43529 2.52190 0.00000 -0.33474 0.98470 -1.78102 1.02215 0.00000 -2.03095 0.23308 1.27559 -1.32335  
 [177] 0.00000 3.61644 0.16339 -0.00896 -0.08075 -0.71200 1.52038 0.00000 -4.69692 -1.59638 -2.23265 2.83118 3.03754 -3.50813 -3.89162 -2.03167  
 [193] -2.21159 -2.85674 0.00000 -2.21245 -1.61433 0.00000 -0.36019 0.00000 -2.66508 0.00000 0.80399 3.09754 0.00000 -1.90830 1.60749 1.11821  
 [209] 0.93676 -0.43651 6.82920 -2.66508 -0.22147 2.22686 -0.54082 -1.78308 -1.59427 2.47131 0.17980 0.00000 -0.01182 1.80786 -0.79549 1.50729  
 [225] 0.00000 0.94068 0.00000 0.34608 -1.56676 -2.25865 -0.86326 -1.26133 -2.30340 0.96516 2.90459 -1.07228 0.00000 -0.12409 -0.55129 0.00000  
 [241] -1.58783 2.23641 3.23785 -1.30345 0.57847 -0.05933 0.42432 0.37040 -0.91460 0.65472 0.19336 0.00128 -2.39695 -1.42305 -0.03788 -4.74451  
 [257] 2.63676 -1.75414 0.69281 0.03081 0.25550 -0.82932 -1.25865 1.43533 -3.77755 -0.29521 -1.77912 0.12859 -3.18843 -5.44879 -2.54321 -3.93199  
 [273] -2.08021 0.00000 0.00000 -2.65267 -1.23374 1.93147 -1.81985 -0.32129 1.93269 -3.15758 -2.71051 2.56480 -2.18170 1.11326 -0.02330 -3.39912  
 [289] 1.40532 0.44222 -1.80706 -0.12513 1.98467 1.47596 2.52197 -1.82031 -3.72060 -1.22684 -2.69378 0.00000 -2.17020 1.50448 0.39797 -2.32648  
 [305] -1.71855 -1.21708 -0.75425 2.17832 -2.11257 -3.28770 -1.93060 1.24481 -1.68643 -0.38308 0.00000 2.48455 0.13753 -0.08639 0.00000 -1.21580  
 [321] 1.72518 0.00000 0.00000 -3.36327 3.29091 0.82858 -1.44264 -2.02209 -0.30059 -2.50859 -0.50946 0.00000 1.08069 -0.34495 -2.47567 -2.18288  
 [337] -2.56479 -0.57093 4.22755 -1.15423 0.43903 1.30543 1.48630 0.00000 -0.76266 -0.16636 -0.74258 0.47345 -3.29622 1.98485 0.00000 3.26546  
 [353] -2.29833 0.50324 0.96533 -3.11029 1.41861 -2.09744 -1.85282 -0.89987 -2.77163 -2.45094 -0.65078 3.02283 -1.80936 -0.91706 -4.80176 -2.35796

