UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ZOOTECNIA E ENGENHARIA DE ALIMENTOS

BÁRBARA DA CONCEIÇÃO ABREU SILVA

Estudo da multicolinearidade em bovinos compostos multirraciais

Pirassununga

2020

BÁRBARA DA CONCEIÇÃO ABREU SILVA

Estudo da multicolinearidade em bovinos compostos multirraciais

Versão Corrigida

Tese apresentada à Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Doutora em Ciências do programa de Pós-Graduação em Zootecnia.

Área de Concentração: Melhoramento Genético Animal

Orientador: Prof. Dr. Joanir Pereira Eler

Pirassununga
2020

Ficha catalográfica elaborada pelo
Serviço de Biblioteca e Informação, FZEA/USP,
com os dados fornecidos pelo(a) autor(a)

162e Abreu Silva, Bárbara da Conceição
 Estudo da multicolinearidade em bovinos
 compostos multirraciais / Bárbara da Conceição
 Abreu Silva ; orientador Joanir Pereira Eler. --
 Pirassununga, 2020.
 104 f.

 Tese (Doutorado - Programa de Pós-Graduação em
 Zootecnia) -- Faculdade de Zootecnia e Engenharia
 de Alimentos, Universidade de São Paulo.

 1. Regressão de Cumeeira. 2. Genética
 Quantitativa. 3. Avaliação Genética. 4. Melhoramento
 Genético. 5. Modelagem. I. Pereira Eler, Joanir,
 orient. II. Título.

DISPENSA DE ANÁLISE ÉTICA

Comunicamos que o projeto de pesquisa abaixo identificado **está dispensado da análise ética** por utilizar animais oriundos de coleções biológicas formadas anteriormente ao ano de 2008, ano da promulgação da [Lei nº 11.794/2008](#) – lei que estabelece procedimentos para o uso científico de animais.

Ressaltamos que atividades realizadas na vigência da referida lei, ou que resulte em incremento do acervo biológico, devem ser submetidas à análise desta CEUA conforme disposto pelo *Conselho Nacional de Controle de Experimentação Animal* (CONCEA).

Projeto: Estudo da multicoliniaridade em bovinos compostos multirraciais

Finalidade: Pesquisa Acadêmica – Tese de Doutorado

Pesquisador: Bárbara da Conceição Abreu Silva

Pesquisador responsável: Prof. Dr. Joanir Pereira Eler (Dep. ZMV)

Instituição: Faculdade de Zootecnia e Engenharia de Alimentos

Informações adicionais: banco de dados comercial pertencente ao Programa Montana Composto Tropical® gerenciado pelo Grupo de Melhoramento Animal e Biotecnologia.

Pirassununga, 1º de outubro de 2019.

Profa. Dra. DANIELE DOS SANTOS MARTINS
Presidente da Comissão de Ética no Uso de Animais

Profa. Dra. CRISTIANE GONÇALVES TITTO
Vice-Presidente da Comissão de Ética no Uso de Animais

BÁRBARA DA CONCEIÇÃO ABREU SILVA

Estudo da multicolinearidade em bovinos compostos multirraciais

Tese apresentada à Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Doutora em Ciências do programa de pós-graduação em Zootecnia.

Data de aprovação: 27 de maio de 2020

Banca examinadora:

Prof. Dr. Joanir Pereira Eler
Faculdade de Zootecnia e Engenharia de Alimentos – Universidade de São Paulo
Presidente da Banca Examinadora

Profa. Dra. Juliana Petrini
Universidade Federal de Alfenas

Prof. Dr. Josineudson Augusto II Silva de Vasconcelos
Faculdade de Medicina Veterinária e Zootecnia - Universidade Estadual Paulista

Profa. Dra. Rachel Bueno Santos de Carvalho
Faculdade de Zootecnia e Engenharia de Alimentos – Universidade de São Paulo

Dr. Rafael Espigolan
Faculdade de Zootecnia e Engenharia de Alimentos – Universidade de São Paulo

DEDICATÓRIA

Dedico este trabalho aos meus amados pais José Dalvo e Josefa Helenice, meus exemplos de vida, dedicação e amor, responsáveis por me ensinar coisas fundamentais para minha personalidade, como o respeito aos mais velhos, a ter força nos piores momentos, o amor e a devoção a Deus. Dedico também a minha querida avó D. Maria de Lourdes (*in memoriam*) e minha tia Maria Auxiliadora (*in memoriam*) que durante sua jornada aqui no plano terrestre foram meus guias e exemplos de mulheres de força, e com certeza aonde estiverem estão extremamente felizes por mais esta conquista em minha vida.

AGRADECIMENTOS

A Deus.

Ao Prof. Joanir Pereira Eler, primeiramente por me aceitar como orientada lá em 2013. Ele confiou em mim e me deu a oportunidade de evoluir profissionalmente e pessoalmente, obrigada também pela grande amizade, por todos os ensinamentos transmitidos, enfim por toda confiança depositada na realização deste trabalho, muito obrigada.

Ao meu companheiro e amor da minha vida, Bruno, pela paciência, cumplicidade e carinho em todos estes anos e em todos os momentos.

À CAPES, pelo suporte financeiro para a execução deste trabalho.

A equipe do Programa Montana Composto pela cessão dos dados.

Ao Prof. Dr. José Bento Sterman Ferraz pela grande ajuda profissional e acadêmica e amizade.

Ao Prof. Dr. Júlio César de Carvalho Balieiro, pelos conselhos e experiências compartilhadas que levarei na minha vida profissional para sempre. Ao Prof. Dr. César Gonçalves de Lima “Cesinha” pela simpatia e grande ajuda profissional e acadêmica, além da amizade. A Prof^a Rachel, sempre companheira, atenciosa dona de uma simpatia única, obrigada por tudo professora a senhora é muito especial.

Aos meus sogros Simone e Junior pelo acolhimento, atenção e carinho por mim, a minha cunhada, Fernanda, pelo carinho e pelas risadas ofertadas. E a toda a família do Bruno, que se tornaram meu segundo lar.

A amiga Elisângela “Li”, pelo carinho, amizade, paciência, simpatia e atenção em todos os momentos, afinal o que seríamos de nós do GMAB sem ela? Obrigada por tudo, Li, você é sensacional!

E durante toda essa trajetória, agradeço aos amigos que passaram todos esses anos pelo GMAB, Isabela (grande parceira), Mikaele “Mika” (pelo apoio sincero

desde a graduação), Gérson Jr (Gersin), Leydiane (Leydinha), Yuri (Yureco), Francisco “Piu-Piu”, Gerardo (Mamani Mamani Mamani), Pâmela (Pam), Lenin, David (Prezadinho), Priscila, Miguel, Jane, Alessandra, Hélio, Rosiane, que estiveram ao meu lado, sempre discutindo ideias e trabalhando em equipe. Aos laços de amizade fortalecidos no doutorado dos amigos: Lais (Laisitaaa), Fernanda (minha Véinha), Hugo (pãozin de queijim), Alisson (Chiquita), Carol, Gabriel, a “Tia Vera” do GMA, ao Rafael Espigolan e a Mirele Poletti. Desejo muita saúde e força e tenho certeza que haverá muito sucesso em suas vidas, pois, acredito no potencial de cada um de vocês.

Ao Bussiman, que apesar de toda a ignorância de uma mula chucra é sem dúvidas peça chave em toda essa tese, obrigada pela cumplicidade “migo”, seu coração é tão grande quanto a sua inteligência. O mundo é seu mocinho! Obrigada também por me deixar roubar a sua vó e a sua mãe de vez em quando!

Ao protótipo de japonês mais convencido e ser humano incrível; Eguti, Felipe (2020) obrigada pela parceria e pela amizade sincera e o melhor trabalho de GWAS para crescimento ponderal padrão internacional da última década! Você é demais!

Aos amigos Madalena (‘Madá’, assim como Mika, verdadeiras amigas desde a graduação), Karen, Roberta, aos casais que são meus grandes amigos, parceiros em todas as horas: Julia e Wagner (Waguinho), Simone (Si) e Rafael (Rafa) – que me deram um presente lindo chamado Helena;

A todos meus anjinhos de quatro patas que me ensinaram o sentido do verdadeiro amor, Lessie (*in memorian*), Yuki (*in memorian*), Yumi, Pitty (*in memorian*), Zen e o Zuck.

Aos Professores e Funcionários da FZEA pela oportunidade da realização do curso de pós-graduação.

E a todos que de forma direta e indireta contribuíram para a realização deste trabalho.

Muito Obrigada.

EPÍGRAFE

“Se eu vi mais longe, foi por estar sobre ombros de gigantes.”

Isaac Newton

RESUMO

ABREU SILVA, B. C. **Estudo da multicolinearidade em bovinos compostos multirraciais**. 2020. 105 f. Tese (Doutorado) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2020.

Apesar de um pouco complexa, a formação de animais compostos representa uma alternativa para a solução de diversos problemas encontrados na produção de bovinos de corte no Brasil, dada a diversidade de climas e regiões presentes em nosso país, a criação destes animais compostos auxilia na qualidade dos rebanhos, no que se refere a aspectos produtivos, reprodutivos e de subsistência. Diante de tantas raças, combinações e parâmetros nos modelos de avaliação genética dos animais compostos, surge um problema matemático: a multicolinearidade, ela ocorre quando as variáveis independentes possuem uma alta correlação, levando então a um confundimento dos estimadores dos coeficientes de regressão. Os principais objetivos deste estudo são: avaliar a estrutura populacional e a diversidade genética ao longo do processo de criação e seleção do Composto Montana®; detectar a presença de multicolinearidade em características de crescimento; obter estimativas dos efeitos genéticos aditivos direto e materno, dos efeitos genéticos não aditivos, assim como dos efeitos fixos, pelos modelos sem covariáveis (SC), modelo tradicional com as covariáveis (REML), quadrados mínimos (QM), regressão de cumeeira (RC), análise de fatores (AF) e componentes principais (PCA). O diagnóstico para a multicolinearidade para estas características foi suficientemente grande para provar que existe este fenômeno e merece atenção especial nas análises de estimação dos componentes de variância pelo método dos quadrados mínimos e na predição dos valores genéticos para estas características. As correções para a multicolinearidade efetuadas foram eficientes para ajustar o β das covariáveis, para todos os métodos. A AF e PCA apresentaram resultados coerentes com a correção da problemática deste estudo, estas análises tem um potencial elevado para inclusão nas avaliações genéticas de bancos de dados em que há o problema de multicolinearidade. A predição dos valores genéticos dos animais para as características deste estudo foi beneficiada pela análise que consistiu na utilização do fenótipo original, com os componentes ajustados para a análise do QM (QMP). A regressão de cumeeira neste estudo pode não ter trazido grandes benefícios, possivelmente pela empiricidade de estimação do parâmetro “k”. Apesar da diferença observada entre os métodos nas análises de comparação de modelos, como o ajuste do β das covariáveis não interfere na predição dos valores genéticos, recomenda-se a aplicação da correção para a estimação dos componentes de variância e dos parâmetros genéticos, cabe ao pesquisador então um maior detalhamento do seu banco de dados para a escolha do modelo parcimonioso mais adequado.

Palavras-chave: Regressão de cumeeira, Estrutura populacional, Análise de Fatores, Componentes Principais.

ABSTRACT

ABREU SILVA, B. C. **Study of multicollinearity in multibreed composite cattle.** 2020. 105 f. P.h.D. Thesis (Doctoral) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2020.

Although somewhat complex, the formation of composite animals represents an alternative for the solution of several problems found in the production of beef cattle in Brazil, given the diversity of climates and regions present in our country, the creation of these composite animals helps in the quality of herds in terms of productive, reproductive and subsistence aspects. In view of so many races, combinations and parameters in the models of genetic evaluation of composite animals, a mathematical problem arises: multicollinearity, it occurs when the independent variables have a high correlation, leading then to a confusion of the estimators of the regression coefficients. The main objectives of this study are: to evaluate the population structure and genetic diversity throughout the creation and selection process of Composite Montana Tropical® beef cattle. Detect the presence of multicollinearity in growth traits. Obtain estimates of direct and maternal additive genetic effects, non-additive genetic effects, as well as fixed effects, by methods without correction (WC), maximum restricted likelihood (REML), ordinary least squares (OLS), ridge regression (RR), factor analysis (FA) and principal component regression (ACP). The diagnosis for multicollinearity for these traits was large enough to prove that this phenomenon exists and deserves special attention in the analysis of estimation of the components of variance by the method of least squares and in the prediction of genetic values for these traits. The corrections for multicollinearity performed were efficient to adjust the β of the covariates, for all methods. The FA and ACP presented results consistent with the correction of the problem of this study, these analyzes have a high potential for inclusion in the genetic evaluations of databases with a multicollinearity problem. The prediction of the genetic values of the animals for the traits of this study was benefited by the analysis that consisted of the use of the original phenotype, with the components adjusted for the analysis of the OLS (QMP). Ridge regression in this study may not have brought great benefits, possibly due to the empirical estimation of the “k” parameter. As much as the methods showed differences in the model comparison analyzes, as the adjustment of the β of the covariates does not interfere in the prediction of genetic values, it is recommended to apply the correction to estimate the variance components and genetic parameters, it is up to the researcher then gave more details to his database to choose the most appropriate and parsimonious model.

Keywords: Ridge regression, Population structure, Factor analysis, Principal component regression.

SUMÁRIO

1. INTRODUÇÃO	12
1.1 Hipótese	14
1.2 Objetivos	14
1.2.1 <i>Objetivos gerais</i>	14
1.2.2 <i>Objetivos específicos</i>	14
Referências	15
2. REVISÃO DE LITERATURA	17
2.1 Bovinos Compostos e o sistema NABC	17
2.2 Heterose e Heterozigose	19
2.3 Multicolinearidade	21
2.4 Metodologias de regressão	22
Referências	25
3. ESTRUTURA POPULACIONAL E DIVERSIDADE GENÉTICA DE UMA POPULAÇÃO DE BOVINOS COMPOSTOS POR ANÁLISE DE PEDIGREE	28
Resumo	28
Abstract	29
3.1 Introdução	30
3.2 Material e Métodos	31
3.3 Resultados e Discussão	36
3.4 Considerações Finais	47
Referências	48
4. DETECÇÃO E DIAGNÓSTICO DA MULTICOLINEARIDADE EM UMA POPULAÇÃO DE BOVINOS COMPOSTOS MULTIRRACIAIS	51
Resumo	51
Abstract	52
4.1 Introdução	52
4.2 Material e Métodos	54
4.2.1 Banco de dados	54
4.2.1.1 Definição da composição racial, heterozigoses e heterozigose materna	55
4.2.2 Identificação da multicolinearidade	57
4.2.2.1 Análise de Correlação	58

4.2.2.3	Decomposição de variância dos coeficientes de regressão	58
4.2.2.4	Fatores de Inflação de variância (FIV)	59
4.3	Resultados e Discussão	59
4.3.1	Correlações	59
4.4	Considerações Finais	67
	Referências	68
5.	MÉTODOS DE TRATAMENTO PARA A MULTICOLINEARIDADE EM UMA AVALIAÇÃO GENÉTICA DE BOVINOS COMPOSTOS MULTIRRACIAIS	70
	Resumo	70
	Abstract	71
5.1	Introdução	72
5.2	Material e Métodos	73
5.2.1	Banco de dados	73
5.3	Resultados	81
5.4	Considerações Finais	95
APÊNDICE I		98

1. INTRODUÇÃO

Os cruzamentos têm se constituído em uma ferramenta de grande importância para obtenção de melhores índices de produtividade na pecuária de corte. Dentre as estratégias de cruzamento, a formação e utilização de bovinos compostos é uma alternativa viável com uma série de fatores favoráveis. Os bovinos compostos são formados pelo cruzamento de duas ou mais raças de forma a obter animais que retêm muita heterose e possuem boa complementaridade entre raças e uniformidade.

O ponto de partida para um programa de seleção é a avaliação genética de todos os animais. Os resultados da avaliação genética são fundamentais para a tomada de decisão de acasalamento. A primeira etapa de um processo de avaliação genética consiste na estimação dos componentes de (co)variância e dos parâmetros genéticos da população e, nos programas de melhoramento genético de animais cruzados e compostos, deve-se levar em conta a multicolinearidade, um fator que afeta diretamente a precisão das estimativas desses parâmetros genéticos.

A multicolinearidade é um fator relacionado com a presença de fortes relações lineares entre as variáveis explanatórias e que pode levar a uma instabilidade nos coeficientes de regressão, aumentando o erro padrão associado aos componentes, sobretudo quando a metodologia de análise é a dos quadrados mínimos. Em bovinos de corte, existem poucos estudos sobre a multicolinearidade causada pela alta parametrização dos modelos para estimação dos efeitos genéticos em populações cruzadas, sobretudo pela inclusão das composições raciais e das heterozigoses (KINGHORN E VERCOE, 1989; SCHOEMAN, 2002; PIMENTEL et al., 2004; ROSO et al., 2005a,b; DIAS et al., 2011; PETRINI et al., 2012; BERTOLI et al., 2019).

Na presença de multicolinearidade, as estimativas dos componentes de (co)variância podem assumir valores absurdos, ou completamente diferentes do encontrado para as mesmas características em populações semelhantes. Os resultados então se tornam muito sensíveis a mudanças nos bancos de dados, ou modificações na estrutura de parametrização dos modelos (Belsley, 1991). Estimar corretamente os parâmetros genéticos da população constitui-se em

uma condição fundamental para o processo de predição do mérito genético de cada animal.

No capítulo 3 desta tese é apresentado um estudo de estrutura populacional por meio de uma análise do pedigree, pois um dos gargalos de um sistema produtivo de bovinos compostos é a depressão pela endogamia, que pode reverter os ganhos obtidos pela heterose, reduzir também a adaptabilidade dos animais a mudanças ambientais (processo evolutivo) ao longo do tempo. O estudo de estrutura populacional visa explicar e demonstrar a influência do processo de seleção, ao longo do programa de formação e melhoramento de bovinos compostos.

No capítulo 4 são apresentadas algumas formas de obtenção do diagnóstico da multicolinearidade, desde formas empíricas como o coeficiente de correlação linear de Pearson entre as variáveis explicativas do modelo, bem como métodos mais acurados como o número de condição da matriz, que pode determinar a força da multicolinearidade. No mesmo capítulo foi avaliada a técnica de determinação dos fatores de inflação de variância que avalia através do coeficiente de regressão estimado o aumento da variância dos mesmos.

No capítulo 5, após identificada e quantificada a presença de multicolinearidade, alguns modelos estatísticos distintos serão aplicados a fim de avaliar a influência da dependência linear entre as variáveis explanatórias na estimação dos componentes de (co)variância e na predição dos valores genéticos de uma população de bovinos compostos, identificada como Composto Montana®.

1.1 Hipótese

Há multicolinearidade entre as variáveis presentes no modelo de regressão, o que influencia diretamente os erros-padrão dos coeficientes, superestimando ou subestimando os mesmos, impactando diretamente os resultados da avaliação genética.

1.2 Objetivos

1.2.1 *Objetivos gerais*

- Estudar os parâmetros que envolvem a estrutura populacional do Composto Montana®.
- Identificar possíveis dependências entre as variáveis genéticas do modelo e comparar o grau de multicolinearidade na avaliação genética da população.

1.2.2 *Objetivos específicos*

- Por meio de análises envolvendo o pedigree, avaliar a estrutura populacional dos bovinos compostos em estudo, assim como avaliar a diversidade genética ao longo do processo de criação e seleção do Composto Montana®.
- Obter estimativas dos efeitos genéticos aditivos direto e materno, dos efeitos genéticos não aditivos, assim como dos efeitos fixos, pelos métodos sem covariáveis no modelo (SC), com covariáveis no modelo (REML), quadrados mínimos (QM), regressão de cumeieira (RC), análise de fatores (AF) e componentes principais (PCA) e comparar esses métodos.
- Realizar a avaliação genética para a predição do mérito genético dos animais considerando os ajustes para a multicolinearidade.
- Verificar a influência dos métodos de correção para multicolinearidade, na ordenação (“ranking”) dos méritos genéticos (DEP) dos animais e na acurácia de predição genética.

Referências

- A.P.; SMITH, A.F.M. (ed.). **Bayesian Statistics**. Oxford: Clarendon Press, 1992. v.4., p.169-193.
- BELSLEY, D. A. A guide to using the collinearity diagnostics. **Computer Science in Economics and Management**,4, 33-50, 1991.
- BERTOLI, C. D.; NETO, J. B.; MCMANUS, C.; COBUCI, J. A.; CAMPOS, G. S.; PICCOLI, M. L.; ROSO, V. Modelling non-additive genetic effects using ridge regression for an Angus–Nellore crossbred population. **Animal Production Science**, v. 59, n. 5, p. 823, 2019.
- CARDOSO, F. F., G. J. M. ROSA, AND R. J. TEMPELMAN. Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. **Journal of Animal Science** 83:1766-1779. 2015.
- CARDOSO, V., QUEIROZ, S.A. AND FRIES, L.A. Estimativas de efeitos genotípicos sobre os desempenhos pré e pós- desmama de populações Hereford x Nelore. **Revista Brasileira de Zootecnia**. 37, 1763–1773.2008.
- DIAS, R. A. P. et al. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. **Livestock Science**, Amsterdam, v. 142, n. 13,p.188–194,2011.
- ELER, J. P. **Teorias e métodos em melhoramento genético animal :sistemas de acasalamento**. Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo.129p.2017.
- FERRAZ, J.B.S.; ELER, J.P. Desenvolvimento de bovinos de corte compostos no Brasil: o desafio do projeto Montana Tropical. IN: Simpósio Nacional De Melhoramento Animal. 2000. **Anais ... Sociedade Brasileira de Melhoramento Animal**, 2000.
- FERREIRA, D. F. **Estatística Multivariada**. 2.ed. Lavras: Ed. UFLA, 2011. 675p.
- GEWEKE, J. Evaluating the accuracy of samplingbased approaches to calculating posterior moments. In: BERNARDO, J.M.; BERGER, J.O.; DAVID, HONGYU, K., SANDANIELO, V. L. M., JUNIOR, G. J. O. Análise de Componentes Principais: resumo teórico, aplicação e interpretação.**E&S - Engineering and Science**.5:1.2016.
- HONGYU, K. **Comparação do GGEbiplot ponderado e AMMI-ponderado com outros modelos de interação genótipo x ambiente**. 2015. 155p. Tese

(Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2015.

KINGHORN, B.P.; VERCOE, P.E. The effects of using the wrong genetic model to predict the merit of crossbred genotypes. **Animal Production**, v.49, p.209-216, 1989.

MISZTAL, I. Old page for BLUPF90 family of programs. Disponível em: <<http://nce.ads.uga.edu/~ignacy/newprograms.html>>. 2014.

PETRINI, J.; DIAS, R.P.; FERNANDA, S.; PERTILE, N.; ELER, J.P.; FERRAZ, J.B.S.; MOURÃO, G.B. Degree of multicollinearity and variables involved in linear dependence in additive - dominant models. **Pesquisa Agropecuária Brasileira**. 47:1743–1750. 2012.

R Core Team (2017). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria.

REGAZZI, A.J. **Análise multivariada, notas de aula INF 766**. Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.

ROSO, V. M., F. S. SCHENKEL, S. P. MILLER, AND J. W. WILTON. Additive, dominance, and epistatic loss effects on preweaning weight gain of crossbred beef cattle from different Bos taurus breeds. **Journal of Animal Science** 83:1780-1787. 2005a.

ROSO, V. M., F. S. SCHENKEL, S. P. MILLER, AND L. R. SCHAEFFER. Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation1. **Journal of Animal Science** 83:1788-1800. 2005b.

SCHOEMAN, S. J., AZIZ, M. A.; JORDAAN, G. F. The influence of multicollinearity on crossbreeding parameter estimates for weaning weight in beef cattle. **South African Journal of Animal Science**. 32:239 – 246. 2002.

2. REVISÃO DE LITERATURA

2.1 Bovinos Compostos e o sistema NABC

A ciência de desenvolvimento dos animais compostos é relativamente recente, tendo surgido no final da década de 1980. Vários grupos estudaram simultaneamente este sistema de acasalamentos, com destaque para o *U.S. Meat Animal Research Center* (USMARC), localizado em Clay Center, Nebraska, Estados Unidos (KOCH et al., 1989; CUNDIFF et al., 1993, 1997, 1998). Basicamente os animais compostos são formados pelo cruzamento de animais de duas ou mais raças, gerando heterose e complementaridade.

Bourdon (1999), explicou que a ciência dos animais compostos deve seguir um planejamento cuidadoso que vai além de apenas cruzamento de várias raças distintas. Os programas de cruzamento devem levar em consideração o mérito genético dos animais que serão utilizados como raças formadoras e o nível de heterose retida deve ser monitorado para garantir que não haja perdas significativas por recombinação gênica. Este valor vai depender majoritariamente do número de raças e da proporção de contribuição de cada uma. Ao incluir diversas raças num sistema de cruzamento, aspectos como simplicidade de manejo e custos envolvidos também são fatores de extrema importância num programa de formação de bovinos compostos.

A complementaridade entre as raças é outro atributo a ser monitorado, uma base genética ampla deve ser utilizada para evitar problemas causados pela endogamia. Outro aspecto a ser observado é a uniformidade do desempenho das progênies, associada diretamente com a manutenção da variação da composição racial ao longo das gerações. E não menos importante é a acurácia da predição genética. Deve-se, portanto, utilizar animais de alto mérito genético, mas não esquecer da acurácia nas decisões de seleção.

Em um cenário tropical brasileiro e buscando incorporar em uma única raça, características de produção de carne e requisitos de adaptabilidade, qualidade de carne, precocidade sexual, entre outras, em 1994, a Agropecuária CFM Ltda e a *Leachman Cattle Co.* dos Estados Unidos se uniram e decidiram criar um composto que atendesse aos requisitos acima. Atualmente o Projeto Montana engloba propriedades localizadas no Brasil, Paraguai e Uruguai. Cerca

de 35 mil vacas são utilizadas como matrizes do programa de melhoramento e desenvolvimento do Montana.

Para atender a um delineamento experimental adequado, principalmente em relação ao número de raças e ao número de animais dentro de cada raça, optou-se pela utilização, na análise genética, do conceito de tipos biológicos, em que as raças foram agrupadas de acordo com essa condição. Os animais formadores da população Montana foram agrupados em quatro tipos biológicos e em um sistema denominado NABC, em que:

- **Grupo N:** representado pelos animais *Bos taurus indicus*. Estes animais contribuem para o programa com as principais características: rusticidade, resistência a parasitas e rendimento de carcaça. As raças principais utilizadas são as vacas Nelore, porém outros representantes deste grupo são as raças Guzerá, Brahman, Gir, Tabapuã entre outras.

- **Grupo A:** nesta categoria são aceitos os animais adaptados aos trópicos de origem européia *Bos taurus taurus*. São responsáveis pelo fornecimento de características muito importantes como: alto grau de adaptabilidade ao clima tropical e bons índices de fertilidade, além de outros aspectos relacionadas à qualidade de carne. As principais raças encontradas neste grupo no programa são: Afrikander, Belmont Red, Bonsmara, Caracu, Romo Sinuano, Senepol e outras.

- **Grupo B:** neste grupo se encontram os animais *Bos taurus taurus* de origem britânica, que contribuem com sua precocidade sexual e de acabamento, além de características de conformação, qualidade de carcaça, carne e crescimento. Como principais raças são encontrados animais: Red Angus e South Devon.

- **Grupo C:** se encontram os animais taurinos *Bos taurus taurus* de origem europeia continental. Animais deste grupo são responsáveis por apresentar elevado potencial de crescimento, rendimento e qualidade de carcaça. Porém um constante controle envolvendo os índices de precocidade sexual e de crescimento são realizados constantemente pois os animais deste grupo são considerados mais tardios. Como raças principais podemos citar: Simental, Gelbvieh, Pardo-Suíço, Charolês, Limousin, Marchigiana.

Além disso, para que um animal seja considerado Montana® ele deve atender a alguns requisitos no que se refere a sua composição racial, conforme demonstrado na Tabela 1.

Tabela 1. Composição racial dos animais Montana®

Grupo racial	Mínimo	Máximo
Número de raças	3	Sem limite
Grupo N	0	6/16
Grupo A	2/16	14/16
Grupo N + Grupo A	4/16	16/16
Grupo B	0	12/16
Grupo C	0	12/16
Grupo B + Grupo C	0	12/16

Fonte: Sumário 2017

Apesar de um pouco complexa, a formação de animais compostos representa uma alternativa para a solução de alguns problemas encontrados na produção de bovinos de corte no Brasil, dada a diversidade de climas e regiões presentes em nosso país, a criação destes animais compostos auxilia na qualidade dos rebanhos, no que se refere a aspectos produtivos, reprodutivos e de adaptação a condições adversas. Outro aspecto positivo na utilização dos compostos, é a obtenção e manutenção de altos níveis de heterose, além da complementaridade e uniformidade nos rebanhos.

2.2 Heterose e Heterozigose

Entender o conceito de heterozigose e de heterose é muito importante na avaliação genética e o nível de heterose esperada depende de diversos fatores como o número e a proporção de raças utilizadas na formação do composto. A heterozigose pode ser definida como um conceito puramente matemático que determina a porcentagem de combinações cromossômicas cruzadas em relação ao total dessas combinações (ELER,2017), ou seja, a heterozigose representa a probabilidade dos alelos de um determinado “locus” provirem de raças distintas (FRIES,1994).

A heterose representa o fenômeno real que é proporcional à heterozigose, ou seja, para 100% de heterozigose, tem-se a heterose máxima. Essa

heterozigose é, no entanto, obtida por meio do acasalamento entre animais de diferentes raças ou linhagens, utilizando um delineamento conhecido como “cruzamento em Dialelo”, no qual os machos de uma raça são acasalados com fêmeas de todas as raças envolvidas no delineamento e, da mesma forma, fêmeas de uma raça são acasaladas com machos de todas as raças envolvidas. A heterose é, então, definida como a superioridade da média das progênies cruzadas em relação a média das progênies das raças parentais envolvidas no cruzamento (ELER,2017).

A equação para o cálculo da heterozigose foi descrita por Eler (2017):

$$HTZ = \sum_{i \neq j} r_i r_j$$

em que, r_i é a proporção do i-ésimo tipo biológico paterno, r_j é a proporção do j-ésimo tipo biológico materno.

Como consequências genéticas dos cruzamentos podemos citar o aumento: da heterozigose, do valor da combinação gênica, do valor genotípico e da média da população. Embora os cruzamentos levem ao aumento da frequência dos heterozigotos na população, a verdadeira causa do aumento do valor genotípico médio e, consequentemente, aumento da média fenotípica dos indivíduos cruzados em relação à média dos produtos das raças parentais, é o aumento do valor da combinação gênica, sendo, portanto, a verdadeira causa da heterose (ELER,2017).

Em bancos de dados de populações de bovinos multirraciais, uma das metodologias mais aplicadas para a obtenção das estimativas dos valores de heterose, é a inclusão nos modelos de análise de regressão, dos efeitos das proporções das composições raciais de cada animal ou a heterozigose de cada composição, como covariáveis nos modelos (VERGARA et al, 2009; BUENO et al., 2011). Outra forma de inclusão desses efeitos pode ser sob a forma de pré-ajuste dos dados, sob diferentes metodologias para a posterior avaliação genética dos animais (MOURÃO et al. 2007; BOCCHI et al. 2008; DIAS et al. 2011).

2.3 Multicolinearidade

Uma relação linear perfeita entre duas ou mais variáveis explanatórias de um modelo de regressão foi definida pela primeira vez como multicolinearidade por Frisch (1934) em um estudo de econometria. Em análises de regressão múltipla, como as avaliações genéticas por exemplo, quando incluímos variáveis explicativas geralmente há uma melhora nos modelos de regressão, porém quando adicionamos variáveis multicolineares podemos ter efeitos prejudiciais na obtenção das estimativas dos parâmetros destes modelos.

A multicolinearidade pode ser ocasionada por vários fatores, dentre eles o tamanho amostral pequeno, ou ainda, quando o número de variáveis explanatórias nos modelos é superior a quantidade de variáveis explicativas, quando há uma alta correlação ($r > 0.90$) entre as variáveis preditoras, estrutura do banco de dados, efeitos redundantes nos modelos de predição. Existe indicação de multicolinearidade na inclusão de diversos efeitos genéticos fixos, como covariáveis, quando há um aumento nos coeficientes associado a um grande erro padrão, ou ainda quando os sinais esperados dos mesmos não atendem as expectativas, temos algumas indicações de multicolinearidade. Alguns autores, trabalhando na presença de multicolinearidade em seus dados associam essas mudanças em virtude da adição ou retirada de uma variável do modelo, destacando mais uma vez a multicolinearidade como causador de tais alterações na obtenção das estimativas dos parâmetros (PIMENTEL, 2004).

Segundo Cardoso et al. (2008), para o sucesso num programa de seleção e de cruzamento, um passo de extrema importância é a obtenção das estimativas confiáveis dos parâmetros genéticos, em que o principal foco é aproveitar o máximo dos efeitos não aditivos entre as diferentes raças. Na avaliação genética de bovinos multirraciais, a colinearidade existente entre as variáveis preditoras dos efeitos fixos tem sido alvo de muitos estudos (Schoeman et al., 2002; Pimentel, 2004; Roso et al., 2005b; Dias, et al., 2011; Petrini et al., 2012; Bertoli, 2016).

2.4 Metodologias de regressão

Alguns métodos são encontrados na literatura para a estimação dos componentes de (co)variância na presença ou não de multicolinearidade, os que serão destacados nesse estudo são: o método dos quadrados mínimos ordinários (QM), a regressão de cumeieira (RC) e a análise de componentes principais (PCA).

O método dos quadrados mínimos ordinários também é chamado de Ordinary Least Square (OLS) quando se considera que os resíduos não são correlacionados e há homocedasticidade. Tendo em vista o modelo linear:

$$y = X\beta + e$$

Em que, y é o vetor de observações,

X é a matriz de incidência dos efeitos fixos,

β é o vetor dos efeitos fixos e

e é o vetor de erros aleatórios,

O estimador dos quadrados mínimos ordinários (OLS) é obtido por:

$$\hat{\beta} = (X'X)^{-1}X'y$$

As covariáveis dos modelos foram agrupadas em uma matriz C , assim, apesar do estimador de mínimos quadrados de $\hat{\beta}$ ser o estimador linear não viesado de mínima variância, seu quadrado médio residual ainda será grande se existirem colinearidades entre as variáveis explanatórias (DIAS,2008).

Outra estratégia utilizada para a redução da multicolinearidade é através da análise de componentes principais. Hongyu et al. (2016), destacam que as principais vantagens da PCA consistem em retirar a multicolinearidade das variáveis, pois esta metodologia de análise permite transformar um conjunto de variáveis originais inter correlacionadas em um novo conjunto de variáveis não correlacionadas (componentes principais). Corroborando com os trabalhos de Hongyu (2015) e Regazzi et al. (2000), que confirmam a redução de muitas variáveis a novos eixos perpendiculares ortogonais que explicam a variação dos dados de forma decrescente e independente.

Na regressão de componentes principais (PCA), transformamos as variáveis independentes em seus componentes principais:

$$C'C = PDP' = Z'Z$$

Em que $\mathbf{D}$ representa uma matriz diagonal dos autovalores de $\mathbf{C}'\mathbf{C}$, $\mathbf{P}$ é o autovetor da matriz $\mathbf{C}'\mathbf{C}$ e $\mathbf{Z}$ é a matriz dos dados (com estrutura similar a $\mathbf{C}$) constituída pelos principais componentes. $\mathbf{P}$ é ortogonal de modo que $\mathbf{P}'\mathbf{P} = \mathbf{I}$. Novas variáveis $\mathbf{Z}$ serão criadas, como médias ponderadas das variáveis originais, como essas novas variáveis são componentes principais, suas correlações entre si são todas zero.

Spearman (1904) em um estudo de psicologia, buscava encontrar um fator comum para escalas cognitivas, como, vocabulário, comunicação, arte e etc., esta metodologia foi muito utilizada em outros estudos da área, porém o método de aplicação ainda era um tanto quanto empírico. Em meados da década de 40, Lawley (1942) fundamentou a técnica em modelos estatísticos e matemáticos, o que difundiu então a possibilidade de aplicação em outras áreas de pesquisa. A análise de fatores (AF) se assemelha em muitos aspectos a PCA, ambas técnicas multivariadas buscam a redução das variáveis em um conjunto de componentes que traduzam a maior parte da variabilidade dos dados por meio de transformações lineares e modelagem das matrizes de covariâncias.

O principal objetivo da AF é encontrar um conjunto de fatores que formem uma combinação linear das variáveis originais ou da matriz de correlações, se houver a presença de alta correlação entre as variáveis as mesmas serão combinadas para formar um fator, esse processo se repete para todas as variáveis incluídas na análise. No estudo em questão a AF foi aplicada na matriz das covariáveis ($\mathbf{X}$), para melhorar a interpretação da solução inicial da análise fatorial e retirar a possível correlação entre os fatores foi aplicada a rotação varimax. O conjunto de fatores que explicaram mais de 90% da variância foram substituídos no modelo de análise dos componentes de (co)variância, para a estimação dos principais parâmetros genéticos.

A Regressão de Cumeira (RC) é outra técnica utilizada para analisar dados de regressão múltipla que sofrem de multicolinearidade, ao adicionar um grau de viés às estimativas de regressão, a RC pois causa um decréscimo nas variâncias das estimativas. Com a adição de coeficientes $\mathbf{k} \geq 0$ à diagonal principal da matriz de correlações ($\mathbf{X}'\mathbf{X}$), um viés é adicionado para que o valor estimado das estimativas seja o da população.

$$E(\hat{\beta}) = B$$

Para a estimação da matriz de covariâncias temos:

$$V(\hat{\beta}) = \sigma^2 R^{-1}$$

Assumindo que,

$$V(\hat{b}_i) = r^{jj} = \frac{1}{1-R_j^2}$$

Em que R_j^2 é o valor do coeficiente de determinação, obtido através da regressão de X_j em outras variáveis independentes. Agora a diagonal da matriz dos elementos de correlação recebe um pequeno valor de k

$$\tilde{B} = (R' + kI)^{-1}X'Y$$

Em que k é o parâmetro de cunha “ridge” sendo $k > 0$, e I é a matriz identidade. Roso et al. (2005a, 2005b) determinaram uma metodologia para a identificação do valor das constantes k .

O viés da estimativa é dado por,

$$E(\tilde{B} - B) = [(X'X + kI)^{-1}X'X - I]B$$

e a matriz de (co)variância:

$$V(\tilde{B}) = (X'X + kI)^{-1}X'X(X'X + kI)^{-1}$$

O valor ideal de k vai depender dos coeficientes de regressão verdadeiros (que estão sendo estimados).

Referências

- A.P.; SMITH, A.F.M. (ed.). **Bayesian Statistics**. Oxford: Clarendon Press, 1992. v.4., p.169-193.
- BELSLEY, D. A .1991. **A guide to using the collinearity diagnostics**. Computer Science in Economics and Management,4, 33-50.
- BERTOLI, C. D., BRACCINI, J., ROSO, V. M. Comparing methodologies to estimate fixed genetic effects and to predict genetic values for an Angus x Nellore cattle population. **Journal of Animal Science**, 94(2), 500–513. 2016.
- BOURDON, R. Composites. **Revista Brasileira de Reprodução Animal**. Belo Horizonte, n.2, p.122-127, 1999.
- CARDOSO, F. F., G. J. M. ROSA, AND R. J. TEMPELMAN. Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. **Journal of Animal Science** 83:1766-1779. 2005.
- CARDOSO, V., QUEIROZ, S.A. AND FRIES, L.A. Estimativas de efeitos genotípicos sobre os desempenhos pré e pós- desmama de populações Hereford x Nelore. **Revista Brasileira de Zootecnia**. 37, 1763–1773. 2008.
- CUNDIFF, L. V., GREGORY, K. E., KOCH, R. M. Cicle V of the germplasm evaluation (GPE) program in beef cattle. In: BEEF RESEARCH – PROGRESS REPORT, n. 4, USDA, Agricultural Research Service, p.3-5, 1993.
- CUNDIFF, L. V. et al. Preliminary results from cicle V of the cattle germplasm evaluation program at the Roman L. Hruska U. S. Meat Animal Research Center. In: GER- MPLASM EVALUATION PROGRAM – PROGRESS REPORT, n. 16, USDA, Agricultural Research Service, p.2-11, 1997.
- DIAS, R. A. P. et al. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. **Livestock Science**, Amsterdam, v. 142, n. 13,p.188–194,2011.
- ELER, J. P. 2017. **Teorias e métodos em melhoramento genético animal :sistemas de acasalamento**. Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo.129p.
- FERRAZ, J.B.S.; ELER, J.P. Desenvolvimento de bovinos de corte compostos no Brasil: o desafio do projeto Montana Tropical. IN: Simpósio Nacional De Melhoramento Animal. **Anais ...** Sociedade Brasileira de Melhoramento Animal, 2000.
- FERREIRA, D. F. **Estatística Multivariada**. 2.ed. Lavras: Ed. UFLA, 2011. 675p.
- FRIES, L.A. **Conexão Braford-DEP. Fundamentos técnicos para seleção e cruzamentos**. Porto Alegre: Conexão Braford DEP. Algumas considerações sobre heterozigose e heterose: p. 9-13.1994.

FRISCH, R. **Statistical confluence analysis by means of complete regression systems**. Oslo: University Institute of Economics, 1934. 192 p

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: BERNARDO, J.M.; BERGER, J.O.; DAVID, HONGYU, K., SANDANIELO, V. L. M., JUNIOR, G. J. O. 2016. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. E&S - Engineering and Science .5:1

HONGYU, K. **Comparação do GGEbiplot ponderado e AMMI-ponderado com outros modelos de interação genótipo x ambiente**. 2015. 155p. Tese (Doutorado em Estatística e Experimentação Agronômica)- Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2015.

KINGHORN, B.P.; VERCOE, P.E. The effects of using the wrong genetic model to predict the merit of crossbred genotypes. **Animal Production**, v.49, p.209-216, 1989.

KOCH, R. M., CUNDIFF, L. V., GREGORY, K. E. Beef cattle breed resource utilization. **Brazilian Journal of Genetics**, v.12, p.55-80, 1989.

MISZTAL, I. Old page for BLUPF90 family of programs. Disponível em: <<http://nce.ads.uga.edu/~ignacy/newprograms.html>>. 2014.

MOURÃO, G.B.; FERRAZ, J.B.S.; ELER, J.P.; BALIEIRO, J.C.C.; BUENO, R.S.; MATTOS, E.C.; FIGUEIREDO, L.G.G. Genetic parameters for growth traits of a Brazilian Bos taurus x Bos indicus beef composite. **Genetics and Molecular Research**, Ribeirão Preto, v. 6, n. 4, p. 1190-1200, 2007

PIMENTEL, E. da C. G. **Alternativas para redução de efeitos de multicolinearidade em modelos de avaliação de efeitos genéticos em bovinos de corte**. 2004. Faculdade de Ciências Agrárias e Veterinárias – Unesp, 2004.

PETRINI, J., R. ANTONIO, P. DIAS, S. FERNANDA, N. PERTILE, J. P. ELER, J. BENTO, S.FERRAZ, AND G. B. MOURÃO. 2012. Degree of multicollinearity and variables involved in linear dependence in additive - dominant models. **Pesquisa Agropecuária Brasileira**. 47:1743–1750.

R Core Team (2017). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria.

REGAZZI, A.J. **Análise multivariada, notas de aula INF 766**, Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.

ROSO, V. M., F. S. SCHENKEL, S. P. MILLER, AND J. W. WILTON. 2005a. Additive, dominance, and epistatic loss effects on preweaning weight gain of crossbred beef cattle from different Bos taurus breeds. **Journal of Animal Science** 83:1780-1787.

ROSO, V. M., F. S. SCHENKEL, S. P. MILLER, AND L. R. SCHAEFFER. 2005b. Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation¹. **Journal of Animal Science** 83:1788-1800.

SCHOEMAN, S. J., M. A. AZIZ, AND G. F. JORDAAN. 2002. The influence of multicollinearity on crossbreeding parameter estimates for weaning weight in beef cattle. **South African Journal of Animal Science**. 32:239 – 246.

SPEARMAN, C. 'General intelligence,' objectively determined and measured. **The American Journal of Psychology**, 15(2), 201–293.1904.

VERGARA, O.D.; ELZO, M.A.; CERON-MUÑOZ, M.F.; ARBOLEDA, E.M. Weaning weight and post-weaning gain genetic parameters and genetic trends in a Blanco Orejinegro-Romosinuano-Angus-Zebu multibreed cattle population in Colombia. **Livestock Science**, Amsterdam, v. 124, n1/3, p. 156-162, 2009

3. ESTRUTURA POPULACIONAL E DIVERSIDADE GENÉTICA DE UMA POPULAÇÃO DE BOVINOS COMPOSTOS POR ANÁLISE DE PEDIGREE

Abreu Silva, B.C.¹; Bussiman, F.O.²; Grigoletto, L.¹; Valderrama, A.S.A¹; Eguti, F.C.¹; Mattos, E.C.¹; Ferraz, J.B.S.¹; Eler, J.P.¹

¹ Departamento de Medicina Veterinária; Faculdade de Zootecnia e Engenharia de Alimentos FZEA/USP – Universidade de São Paulo. Av. Duque de Caxias Norte, 225. 13635-900. Pirassununga/SP.

² Departamento de Nutrição e Produção Animal; Faculdade de Medicina Veterinária e Zootecnia FMVZ/USP – Universidade de São Paulo. Av. Duque de Caxias Norte, 225. 13635-900. Pirassununga/SP.

Resumo

O programa para criação do Montana Composto foi desenvolvido para atender a necessidade de animais adaptados aos trópicos que mantêm altos níveis de produtividade, qualidade da carne e da carcaça, resistência a parasitas. O objetivo deste estudo é verificar a estrutura da população de bovinos compostos, por meio de análise do pedigree, bem como inter-relacionar com os ganhos genéticos para as características de crescimento ao longo dos anos de seleção e formação da população. Entender os parâmetros populacionais bem como a estrutura e a variabilidade genética de um rebanho que está em constante processo de seleção para os ganhos em produtividade é imprescindível, pois um dos gargalos de um sistema produtivo de bovinos de corte são as perdas por depressão pela endogamia e a redução da adaptabilidade dos animais a mudanças ambientais (processo evolutivo) ao longo do tempo. O pedigree deste estudo é robusto, porém sua integridade foi um pouco baixa, o pode ser influenciada pela entrada de animais com genealogia desconhecida, provavelmente pela prática da utilização de reprodutores múltiplos nos sistemas de acasalamento. O programa está priorizando a manutenção da diversidade e variabilidade genética na população como visto por meio do tamanho efetivo da população, os níveis de endogamia. A possibilidade de inclusão novas raças no sistema auxiliam na manutenção da heterozigose ao longo das gerações, outro indicativo também de retenção de heterose pode ser visto nas tendências

positivas encontradas para as características de crescimento. Os componentes de (co)variância encontrados indicam que as características são herdáveis e podem ser incluídas em programas de seleção.

Palavras-chave: Parâmetros populacionais; Heterose; Heterozigose; Montana Composto®.

Abstract

Developed to meet the need for animals adapted to the tropics that maintain high levels of productivity, meat and carcass quality, resistance to parasites, the Montana Composto® breeding program started. The aim of this study is to verify the population structure of Montana composite cattle, through analysis of the pedigree, as well as to interrelate with the genetic gains for the growth traits over the years of population selection and formation. Understanding the population parameters as well as the genetic structure and variability of a herd that is in a constant process of selection for gains in productivity is essential, as one of the bottlenecks of a beef cattle production system is the losses due to inbreeding depression and reducing the adaptability of animals to environmental changes (evolutionary process) over time. The pedigree of this study is quite large, but its integrity was a little low, which may be influenced by the entry of animals with unknown genealogy, probably by the practice of using multiple sires in the mating systems. The program is satisfactorily prioritizing the maintenance of genetic diversity and variability in the population as seen through the effective size of the population, the levels of inbreeding, the possibility of including new breeds in the system help in maintaining heterozygosity over the generations, another indication also of heterosis retention can be seen in the positive trends found for the growth traits, the components of (co) variance found indicate that the traits are inheritable and can be included in selection programs.

Keywords: Genealogical parameters; Heterozygosity; Heterosis; Montana Composto®.

3.1 Introdução

Na década de 90 surgiu, por meio de uma *joint venture* entre a Agropecuária CFM Ltda e a *Leachman Cattle Co.* dos Estados Unidos, a idéia de criação de uma raça adaptada aos trópicos que incorporasse características de adaptação e eficiência produtiva, surgiu, assim, o Programa Montana Composto. Para atender a um delineamento experimental adequado, principalmente em relação ao número de raças e ao número de animais dentro de cada raça, optou-se pela utilização, na análise genética, do conceito de tipos biológicos em que as raças foram agrupadas de acordo com essa condição. Os animais formadores da população Montana foram agrupados em quatro tipos biológicos denominação que recebeu a sigla NABC (FERRAZ; ELER; GOLDEN, 1999a, 1999b), em que:

O grupo N é representado pelos animais *Bos taurus indicus*. A principal contribuição deste grupo envolve características como rusticidade, resistência a parasitas e rendimento de carcaça. A raça predominante nesse grupo é a Nelore, porém existem outras raças como Guzerá, Brahman, Gir, Tabapuã. No grupo A se encontram os animais *Bos taurus taurus* adaptados aos trópicos que agregam características como alto grau de adaptabilidade ao clima tropical e bons índices de fertilidade, além de outros aspectos relacionadas à qualidade de carne. Como principais raças encontradas neste grupo podemos citar os bovinos Afrikander, Belmont Red, Bonsmara, Caracu, Romo Sinuano, Senepol e outras.

Os representantes do grupo B são os animais *Bos taurus taurus* de origem britânica, que contribuem com sua precocidade sexual e de acabamento, além de características de conformação, qualidade de carcaça, carne e crescimento. Representados pelas raças Red Angus e South Devon. Finalmente no grupo C, animais *Bos taurus taurus* de origem europeia continental no qual as características exploradas envolvem um elevado potencial de crescimento, rendimento e qualidade de carcaça. Destacam-se os animais das raças Simental, Gelbvieh, Pardo-Suíço, Charolês, Limousin e Marchigiana.

O estudo de estrutura populacional aliado à avaliação do progresso genético dos rebanhos é imprescindível para o desenvolvimento dos programas de seleção e melhoramento genético, determinando a direção das possíveis

mudanças a fim de garantir a variabilidade genética da população selecionada. O objetivo deste estudo foi identificar os principais parâmetros populacionais e sua inter-relação com o progresso genético em características de crescimento ponderal em bovinos compostos.

3.2 Material e Métodos

Por se tratar de um banco de dados comercial e já existente, com animais oriundos de coleções biológicas formadas anteriormente ao ano de 2008, ano da promulgação da Lei nº 11.794/2008 – lei que estabelece procedimentos para o uso científico de animais, este estudo foi dispensado da análise ética.

Os bancos de dados fazem parte do programa de melhoramento genético de bovinos compostos, Montana Composto® e estão sob responsabilidade do Grupo de Melhoramento Animal e Biotecnologia (GMAB) da Universidade de São Paulo. Os animais foram criados em diversas fazendas localizadas em várias regiões do Brasil e também no Uruguai e Paraguai.

O desenvolvimento deste estudo se deu em duas etapas, sendo a primeira responsável pela determinação da estrutura da população durante o processo de formação e seleção do composto. Na segunda etapa, uma amostra da população que continha dados fenotípicos para características de desenvolvimento ponderal, foi utilizada para a estimação dos componentes de (co)variância e cálculo de tendências genéticas a fim de identificar como o processo de seleção tem sido feito nos últimos anos, bem como a sua relação com parâmetros que influenciam diretamente na variabilidade genética e no ganho produtivo.

3.2.1 Banco de pedigree e análise de estrutura populacional

O arquivo de registros do banco de pedigree consistia em informações de 1.309.462 indivíduos, dentre eles 5.737 touros e 471.368 vacas, nascidos entre 1980 e 2018. Para a determinação do grupo genético dos animais foram estabelecidos alguns critérios envolvendo as informações da porcentagem dos tipos biológicos, o que classificou os animais em 5 grupos genéticos (Tabela 1).

Os animais compostos considerados como Montana pertencem aos grupos renumerados de 14 a 18.

Tabela 1 - Critérios de classificação dos animais do banco de pedigree

Grupo Genético	NABC	NABCr ¹	Critério (%)	N ²
Outros	XXXX	0	Qualquer composição que não atenda às condições abaixo	200.566
$\frac{3}{4}$	12XXX	1	60 < N < 90	52.800
	X12XX	2	60 < A < 90	25.669
	XX12X	3	60 < B < 90	13.165
	XXX12	4	60 < C < 90	2.252
Raças Puras	16000	5	N ≥ 90	669.805
	01600	6	A ≥ 90	3.947
	00160	7	B ≥ 90	10.301
	00016	8	C ≥ 90	501
F ₁	8800	9	40 ≤ N ≤ 60 e 40 ≤ A ≤ 60	5.565
	8080	10	40 ≤ N ≤ 60 e 40 ≤ B ≤ 60	67.894
	8008	11	40 ≤ N ≤ 60 e 40 ≤ C ≤ 60	24.678
	0880	12	40 ≤ A ≤ 60 e 40 ≤ B ≤ 60	3.739
	0808	13	40 ≤ A ≤ 60 e 40 ≤ C ≤ 60	99
Montana	4444	14	18,75 < N < 31,25 e 18,75 < A < 31,25 e 18,75 < B < 31,25 e 18,75 < C < 31,25	16.671
	4480	15	18,75 < N < 31,25 e 18,75 < A < 31,25 e 43,75 < B < 56,25	5.666
	4804	16	18,75 < N < 31,25 e 18,75 < A < 31,25 e 18,75 < C < 31,25	22.071
	4840	17	18,75 < N < 31,25 e 43,75 < A < 56,25 e 18,75 < B < 31,25	63.959
	MMMM	18	N ≤ 37,5 e 12,5 ≤ A ≤ 87,5 e B ≤ 75 e C ≤ 75 e (N + A) ≤ 0,25 e (B + C) ≤ 75	1.024

¹ NABC renumerado. ² Número de indivíduos. Fonte: Própria Autoria

3.2.2 “Completeness”

Esta medida demonstra o quão completo e profundo (informativo) é o pedigree em estudo. Dada como $E_q G_i = \sum (1/2)^n$, em que n representa o número de gerações que separa cada indivíduo de cada ancestral conhecido. Para cada geração encontrada no pedigree foi calculada uma média correspondente.

3.2.3 Endogamia

A taxa de endogamia foi calculada de acordo com Falconer & Mackay (1996) por meio da fórmula $\Delta F = 1 - E_q G_i - 1 \sqrt{1 - F_i}$, em que F_i e $E_q G_i$ são os coeficientes de endogamia e o número de gerações completas equivalente para cada indivíduo i , respectivamente.

3.2.4 Intervalo de Gerações

O intervalo de geração foi calculado em toda a população considerando a idade média dos pais no nascimento dos filhos. (FALCONER & MACKAY, 1996).

3.2.5 Tamanho efetivo populacional

O primeiro conceito de tamanho efetivo de uma população (N_e) foi proposto por Wright (1923), em que determinado número de indivíduos de uma população ideal apresentaria a mesma taxa de mudança genética que uma população real, ponderada pela taxa de endogamia média da população.

Para este estudo o N_e foi calculado por cinco métodos: a partir da informação de pedigree, para toda a população, de acordo com as informações censitárias (Cens- Wright, 1923); ponderado pela média da taxa de endogamia dos indivíduos e dos pais a cada t geração (F_p – Falconer & Mackay, 1996); ponderado pela taxa de endogamia dos pais a cada t geração (F_g – Falconer &

Mackay, 1996); ponderado pelo coeficiente da regressão logarítmica do ano de nascimento do animal e o intervalo de geração (Ln – Perez-Enciso, 1995) e através da metodologia que contempla o número de gerações equivalentes completas (Ecg – Gutiérrez et al., 2009). Todas as análises foram realizadas no software PopRep (GROENVELD, 2009).

3.2.6 Parâmetros genealógicos

Os fundadores foram determinados como sendo os animais que continham a informação desconhecida dos seus ancestrais. O número efetivo de fundadores foi calculado de acordo com a metodologia de Lacy (1989) em que $fe = 1/\sum(p_i^2)$ em que p_i é a proporção de genes da população descendente do fundador i . O número mínimo de ancestrais necessários para garantia da diversidade genética (GD) da população de referência constitui-se no número efetivo de ancestrais (fa) calculado como proposto por Boichard et al. (1997).

Para identificar o gargalo da população foi calculada também a razão entre o fe e o fa (Boichard et al., 1997). Outra forma de calcular a GD é o número de fundadores que produziriam uma população com a mesma diversidade de alelos fundadores da população, chamado de número de genoma equivalente do fundador (fge) (Lacy, 1989). As análises foram realizadas no software CFC (SARGOLZAEI et al., 2006).

3.2.7 Estratificação da população

Foi calculada a média do coeficiente de parentesco dos grupos genéticos. Para uma apresentação gráfica dos resultados foi utilizado o pacote heatmap do software R (R Development Core Team, 2016).

3.2.8 Banco de fenótipos e análises genéticas

Para a determinação dos componentes de (co)variância e estimação dos parâmetros genéticos, utilizou-se uma amostra da população original em que foram fenotipados 285.823 animais, 680.551 no pedigree filhos de 4.263 touros

e 350.197 vacas. As características analisadas (Tabela 2) foram peso a desmama (PD, Kg, obtido em torno dos 205 dias), peso ao sobreano (P14, Kg, obtido em torno dos 420 dias) e ganho de peso pós desmama (GP, Kg, obtido pela diferença do peso ao sobreano em relação ao peso à desmama. $GP = (\text{peso ao sobreano} - \text{peso à desmama}) / \text{NDEP} \times 215$. NDEP = número de dias entre pesagens e 215 corresponde ao número de dias entre a desmama (205 dias) e o sobreano (420 dias).

Tabela 2 – Estatística descritiva do banco de fenótipos

CARAC.	N	MÉD	MIN	MÁX	DP	NGC
PD	273.346	197.79	80	320	38.65	3326
GP	104.648	77.13	0.50	229	44.05	1245
P14	113.203	273.49	101	450	55.82	1199

N= número de registros, Méd= média, Min= mínimo, Máx= máximo, DP= desvio padrão, NGC= número de grupo de contemporâneos, NM= número de machos, NF= número de fêmeas. Fonte: Própria Autoria

Os grupos contemporâneos de desmama (GCDES) foram obtidos pela concatenação de fazenda + ano de nascimento (ou safra) + sexo + grupo de manejo à desmama (GMANDES). Os grupos de contemporâneos para o ganho de peso (GCGP) e os de sobreano (GC14) foram obtidos pela concatenação de fazenda + ano de nascimento (ou safra) + fazenda de sobreano + grupo de manejo ao sobreano (GMAN14). Ressalta-se que o grupo de manejo à desmama (GMANDES) não integrou a concatenação para formação dos grupos de contemporâneos ao sobreano (GC14 e GCGP), ou seja, para as características de sobreano, o GMANDES foi incluído, no modelo, como efeito aleatório não correlacionado. As razões para utilização do GMANDES como efeito aleatório não correlacionado são bem descritas em SANTANA et al., (2013) e Pedrosa et al., (2014).

Para a estimação dos valores genéticos, foi utilizado um método bayesiano utilizando o software GIBBS1F90, programa pertencente ao pacote BLUPF90 (MISZTAL et al., 2002). Cada análise consistiu de uma cadeia simples de tamanho de 600.000 amostras, descartando as 100.000 primeiras amostras, salvando num intervalo a cada 100 amostras. Para as inferências finais foi

utilizado a informação de 5.000 amostras. Uma análise tri-característica foi realizada e o modelo estatístico incluiu como efeitos fixos o grupo de contemporâneo, classe da idade da mãe ao parto, idade à pesagem (linear) como covariável. Como efeitos aleatórios foram incluídos os efeitos genéticos direto e materno do animal e efeito de ambiente permanente materno. Para as características GP e P14 foi incluído também o grupo de manejo a desmama como efeito aleatório não correlacionado.

As tendências genéticas foram expressas como a regressão dos valores genéticos para as respectivas características sobre o ano de nascimento dos animais. A base genética adotada consistiu na média das DEPs de todos os animais da população formadora do Montana®, e que nasceram até 2007, ou seja, 10 anos de seleção atrás.

3.3 Resultados e Discussão

3.3.1 Análises de pedigree

O pedigree apresentou um total de 12 gerações totalizando 1.188.251 animais filhos de 5.737 touros e 471.368 matrizes. Os animais Montana foram criados para a sobrevivência em climas tropicais, os animais são criados a pasto e a recria, com alguma suplementação. O processo reprodutivo em sua maioria é por monta natural. A média obtida para o intervalo de gerações da população foi de 5,8 anos, o que está dentro do esperado, dado que há uma grande quantidade de raças distintas envolvidas na formação do composto Montana, principalmente em relação às matrizes. Outras raças taurinas apresentaram intervalos de seleção equivalentes em diversos países, por exemplo de 6 à 6,6 anos em raças africanas (PIENAAR et al.,2015; ABIN et al.,2016), em raças taurinas mexicanas que variaram de 4,2 à 8,5 anos (UTRERA et al.,2018), em raças espanholas 3,7 à 5,5 anos (GUTIERREZ et al.,2003) e em populações zebuínas brasileiras 3,3 à 7,5 anos (CAVANI et al.,2018; AMARAL et al.,2019).

Para os caminhos de seleção entre touro para filho (ss), touro para filha (sd), vaca para filho (ds) e vaca para filha (dd), os intervalos encontrados foram 6,7; 6,5; 5,3 e 5,6 anos respectivamente, o que se assemelhou ao encontrado

por Utrera et al., (2018) com a raça Simmental. Santana Jr. et al., (2012a) também encontraram intervalos de geração para os caminhos de seleção ds e dd menores do que ss e sd, para as raças Bonsmara e Marchigiana. É importante destacar que essas raças também estão presentes na formação do Montana, o que corrobora com os resultados encontrados.

A porcentagem média do *completeness* para os animais nascidos nos últimos 10 anos foi :com 1 geração de profundidade – 48,5; com 2 gerações – 44,2; com 3 gerações- 39,9 ; com 4 gerações – 35,5; com 5 gerações – 30,9 e com 6 gerações – 26,7. A Figura 1, apresenta a porcentagem média de *completeness* até a 12ª geração para toda a população. Surpreendentemente, embora o pedigree seja bastante grande, notou-se que a integridade do pedigree observada foi um pouco baixa, o que pode ter sido influenciado pela entrada de animais com genealogia desconhecida, além disso, como o composto é aberto na formação dos animais é permitida a inclusão de várias raças.

Figura 1 – Porcentagem de *Completeness* por geração.

Fonte: Própria autoria.

Foi encontrado nesse estudo uma baixa taxa de endogamia ΔF (0,01%), o que seria de se esperar em um programa de cruzamentos em que se busca a

retenção de heterose. A endogamia tem sido monitorada desde o início do programa de formação do composto. A endogamia paterna foi maior do que a materna (0,14% e 0,05%, respectivamente), nota-se que o ΔF teve pouca variação ao longo do período estudado (Figura 2). Picolli et al. (2014) encontraram taxas de endogamia médias muito semelhantes às encontradas neste estudo para as raças Angus (0,003%), Devon (0,05%), Hereford (0,02%) e Shorthorn (0,01%).

Os valores de endogamia encontrados para esta população são baixos, o que pode estar relacionado com o fato do Montana se tratar de uma população de animais compostos, outro parâmetro observado foi um baixo coeficiente de ancestralidade (4.20×10^{-6}), este valor representa a contribuição das populações ancestrais para a população atual. Outro fator que corrobora para estes baixos valores encontrados é o rígido controle aplicado no delineamento dos acasalamentos durante o processo de seleção.

Figura 2 – Endogamia dos indivíduos, materna, paterna e taxa de endogamia.

Fonte: Própria autoria.

As estimativas do tamanho efetivo populacional foram de 3768, 951, 440, 845, 339 para os métodos Cens, Fp, Fg, Ln e Ecg respectivamente. Houve uma pequena amplitude nos resultados encontrados e estas estimativas indicam que

a média do N_e dos animais está em torno de 650 animais efetivos. Em consonância com estes resultados, Peripolli et al. (2019) em um estudo genômico de corridas de homozigose com animais Montana®, encontrou um tamanho efetivo populacional, baseado em 50 gerações, de cerca de 528 animais e nas últimas 5 gerações de 128 animais.

A Figura 3 destaca a distribuição do número efetivo populacional por meio das diferentes metodologias ao longo dos anos. A escolha do melhor método para a contagem do tamanho efetivo populacional é relativa e depende de diversos fatores, o método censitário não é o mais indicado para populações que estão sob o regime de seleção. As diferentes metodologias para a estimação do tamanho efetivo populacional indicaram que a população se encontra viável a longo prazo, mantendo a diversidade genética. Historicamente para populações sob conservação genética há uma indicação de N_e mínimo de 50 indivíduos (FAO,1988) e para a manutenção do potencial evolutivo de uma raça, indica-se pelo menos 500 indivíduos constituindo o tamanho efetivo (FRANKHAM,1995).

Figura 3- Distribuição do tamanho efetivo populacional ao longo dos anos.

Fonte: Própria autoria.

O número médio de filhos por touro foi em torno de 82 animais enquanto que o número de filhos por vaca foi de 2,4 animais (Tabela 3). A porcentagem de animais endogâmicos foi de apenas 4,7%. O coeficiente médio de endogamia da população foi de 0,10% enquanto que para o ano de 2018 ele foi de 0,14%. Considerando apenas os animais endogâmicos esses mesmos coeficientes foram 1,99% e 0,78%, respectivamente.

O número de genoma equivalente foi em torno de 46% do número efetivo de fundadores. O número efetivo de não fundadores foi de 1.662,41 animais, como o valor encontrado para f_a foi inferior ao encontrado para o f , pode-se dizer que houve uma pequena perda na diversidade genética da população pelos efeitos de gargalo. Quando os valores são expressos em porcentagem, observa-se que a GD da população em estudo é de 99,94%, com perdas devido à deriva genética de aproximadamente 0,03%. Quando se avalia a perda de diversidade genética acumulada desde a fundação da população encontra-se o total de 0,06% e a perda devido a contribuição desigual dos fundadores também foi em torno de 0,03%.

Tabela 3- Análise de probabilidade de gene origem, diversidade genética, tamanho efetivo e número de subpopulações equivalentes para o pedigree.

Parâmetros	População
Número de Animais	1.309.462
Número de Touros	5.737
Filhos	470.427
Número de Vacas	471.368
Filhos	1.114.181
Número de animais endogâmicos	61.206
Porcentagem de animais endogâmicos	4,7%
Coeficiente médio de endogamia (toda população)	0,1%
Coeficiente médio de endogamia (para os animais endogâmicos)	1,9%
Média do coeficiente de parentesco	0,001
Média de gerações equivalentes completas	1,44
Número de fundadores(f)	194.593
Número efetivo de fundadores (f_e)	1.925,44
Número efetivo de ancestrais (f_a)	972,63
Genoma equivalente dos fundadores(f_{ge})	892,14
Número efetivo de não-fundadores	1.662,41
Razão f_e/f_a	1,98
Razão f_{ge}/f_e	0,46

Fonte: Própria autoria.

As razões f_e/f_a e f_e/f_{ge} indicam a igualdade de contribuição dos animais para o desenvolvimento da população e os valores mais próximos de 1 são os ideais (SANTANA Jr. et al,2012b). A razão f_e/f_a encontrada neste estudo foi menor que 2 (1,98), Boichard et al. (1997) afirmaram que quanto mais altos os valores desta razão, maior é o efeito do “gargalo” na população. Esses valores geralmente são mais elevados em populações em que ocorre o uso intensivo das técnicas reprodutivas, como a inseminação artificial, transferência de embriões. A razão f_{ge}/f_e (0,46), valor considerado baixo, pode ser um indicativo de perda na diversidade genética, devido à deriva genica (LACY, 1989; BOICHARD et al., 1997). A quantidade de informação presente no pedigree e a presença de fundadores com idades mais avançadas, mas, que ainda preservam

relações com animais mais jovens na população de referência pode facilitar o aumento da oscilação genética, em outras palavras, o elevado número de gerações presentes no pedigree ressaltou o fenômeno da deriva.

Subdividindo o pedigree e agrupando a informação apenas dos animais Montana[®], ou seja, dos grupos NABCr 14 a 18 (Tabela 4), nota-se que os grupos 14 e 18 apresentaram o menor número de fundadores que explicam cerca de 50% da variabilidade genética do grupo, o que provavelmente está relacionado com a diversidade genética dentro do grupo. Apesar disso, para todas as composições, os valores de heterozigose encontrados foram de alta magnitude. Este parâmetro, que se associa linearmente com a retenção de heterose, aponta que a utilização de animais compostos oferece diversas oportunidades de ganho nos programas de cruzamento.

Tabela 4- Número de fundadores, endogamia, heterozigose para os grupamentos genéticos dos animais Montana[®]

NABCr	NF	NF50	F	1-F	NFe	Fe	1-Fe
14	10.462	22	0,06	99,94	92	10,57	89,43
15	5.326	30	0,12	99,88	154	4,38	95,62
16	11.559	65	0,01	99,99	45	7,22	92,78
17	22.580	65	0,03	99,97	877	2,16	97,84
18	101	20	0,06	99,94	27	2,31	97,69

NF= Número de fundadores, NF50 = Número de fundadores que explicam 50% da variância, F= coeficiente de endogamia, 1-F= heterozigose, NFe = Número de fundadores endogâmicos, Fe = coeficiente de endogamia dos animais endogâmicos, 1-Fe = heterozigose dos animais endogâmicos. Fonte: Própria autoria.

A amplitude observada tanto nos valores de endogamia quanto nos valores dos parentescos entre os grupos genéticos, foi muito próxima de zero, corroborando com a ciência de formação dos animais compostos em que a endogamia é evitada. Ao encontrar segmentos de homozigose longos em uma amostra da população de animais Montana[®], Peripolli et al., (2019) associaram esses resultados ao desenvolvimento recente da população e à utilização de um painel de baixa densidade para as análises. No entanto, os mesmos autores destacaram que, pela constante inserção de novas raças nos sistemas de

cruzamento, diversas combinações genicas foram encontradas dentro de cada tipo biológico, o que influencia diretamente nos níveis de endogamia da população, levando à redução desses níveis.

O parentesco entre as composições (grupos genéticos) variou de 0 a 0,04% (Figura 4), o parentesco dentro dos grupos foi em média 13,9 vezes maior que o parentesco entre os grupos. Os grupos 14 e 18 são os mais aparentados intragrupo, o que pode ocasionar então uma menor variabilidade genética quando comparado aos demais grupos genéticos do Montana. Por consequência esses grupos apresentaram o menor número de fundadores explicando ao menos 50% da variabilidade genética dentro do grupo.

Figura 4- Parentesco entre os grupos genéticos.

Fonte: Própria autoria

3.3.2 Análises genéticas

As estimativas dos coeficientes de herdabilidade encontradas foram 0,27; 0,11 e 0,30 respectivamente para as características PD, GP e P14 (Tabela 5), valores semelhantes foram encontrados em análise genômica de passo único, com estimativas médias dos coeficientes na magnitude de 0,30; 0,16 e 0,30 PD, GP e P14 respectivamente (GRIGOLETTO et al., 2019). A herdabilidade para essas características PD e P14 aponta um ganho em potencial para a inclusão em programas de melhoramento genético.

A correlação genética (Tabela 5) entre PD e o GP foi negativa e de baixa magnitude (-0,21), este valor negativo, embora relativamente baixo, associado com o valor da correlação ambiental que também foi negativa (-0,26), pode ser um indicativo de interação genótipo x ambiente. Santana Jr. et al. (2012, 2013, 2014), reportaram forte influência da interação genótipo ambiente na mesma população em estudo, por meio de diversas metodologias (normas de reação, análise multicaracterística, análise de cluster), encontrando a existência de variação entre as diferentes regiões geoclimáticas, o que interfere diretamente nas estimativas das correlações. Modelos que consideraram a interação genótipo ambiente apresentaram valores superiores, reforçando a importância inclusão em posteriores modelos de predição para os animais em estudo.

No período compreendido entre a desmama e a pós-desmama os animais passam por situações de estresse alimentar muito severo, principalmente em condições tropicais de manejo a pasto, e a compensação desse crescimento se dá, então, no pós-desmama, por meio da alimentação de melhor qualidade. Outra hipótese seria em relação aos genes que determinam as duas características. Possivelmente os genes que controlam cada uma das características poderia se encontrar em regiões diferentes no genoma (SANTANA et al., 2012). Cabe ressaltar, ainda, que, nesta população analisada, a taxa de descarte à desmama é muito alta (40 a 50%), o que pode ser observado na diferença significativa (cerca de 160.143 bezerros) dos animais fenotipados para as duas características (Tabela 2).

A correlação do PD e P14 (0,89) foi positiva e de alta magnitude, valores similares foram encontrados (0,87) por Grigoletto et al. (2019), indicando que os genes que regulam a característica são os mesmos ou estão em regiões muito

próximas. Bezerros que apresentam peso a desmama maior tendem a apresentar peso ao sobreano superior também. Mais uma vez, a alta taxa de descarte à desmama pode estar impactando esta correlação, uma vez que animais com baixo peso à desmama não são pesados ao sobreano. A herdabilidade para o PD materno, neste estudo, foi muito baixa ($0,02 \pm 0,01$). O grupo de manejo à desmama, dado como proporção da variância fenotípica, para o P12 foi de $8,96 \pm 0,53\%$ e para o GP $12,96 \pm 0,71\%$. Ou seja, os lotes de desmama (GMANDES) explicam, em boa proporção, a variância dessas duas características.

Tabela 5- Estimativas do coeficiente de herdabilidade na diagonal, correlação genética acima da diagonal e correlação residual abaixo da diagonal.

	PD	GP	P14
PD	$0,27 \pm 0,01$	$-0,21 \pm 0,04$	$0,89 \pm 0,01$
GP	$-0,26 \pm 0,01$	$0,11 \pm 0,01$	$0,25 \pm 0,04$
P14	$0,55 \pm 0,01$	$0,66 \pm 0,01$	$0,30 \pm 0,01$

Fonte: Própria autoria.

Os gráficos de tendência genética (Figura 5) apontaram para ganhos em todas as características observadas durante o período de seleção. A tendência média foi de aproximadamente 0,44 Kg/ano para o PD, aproximadamente 0,15Kg/ano para o GP e de 0,59 Kg/ano para o P14. Os ganhos em heterozigose anuais foram em torno de 0,02% enquanto que a endogamia caiu na magnitude de -0.0004% ao ano, confirmando a ciência proposta pelos compostos de aumento da heterozigose nos rebanhos, a utilização de diversas raças e o manejo dos cruzamentos potencializa o aproveitamento da heterose.

Figura 5- Tendências genéticas, endogamia e heterozigose para as características de crescimento.

Fonte: Própria autoria.

3.4 Considerações Finais

O pedigree é bastante grande, porém sua integridade foi um pouco baixa, o que pode, pelo menos em parte, estar relacionado com a utilização de reprodutores múltiplos (grupos de RM) em que parte dos animais da população analisada não tem pai conhecido. Em contrapartida, o programa está notoriamente priorizando a manutenção da diversidade e variabilidade genética na população como visto por meio do tamanho efetivo da população, os níveis de endogamia encontrados neste estudo foram muito baixos, não apresentando riscos para problemas como a depressão pela endogamia. A possibilidade de inclusão de novas raças no sistema contribui para a manutenção da heterozigose e, conseqüentemente, da heterose ao longo das gerações. Outro indicativo de retenção de heterose pode ser visto nas tendências positivas encontradas para as características de crescimento. As características em estudo apresentam coeficientes de herdabilidade suficientes para apresentarem resposta à seleção. O coeficiente de herdabilidade para GP, no entanto, indica que a resposta à seleção pode ser bem menor do que o esperado para PD e P14.

Referências

- ABIN, S.; THERON, H. E.; VAN MARLE-KÖSTER, E. Population structure and genetic trends for indigenous African beef cattle breeds in South Africa. **South African Journal of Animal Sciences**, v. 46, n. 2, p. 152–156, 2016.
- BOICHARD, D.; Maignel, L.; VERRIER, É. The value of using probabilities of gene origin to measure genetic variability in a population. **Genetics Selection Evolution**, v. 29, n. 1, p. 5, 1997. Disponível em: <<http://www.gsejournal.org/content/29/1/5>>.
- CAVANI, L.; SILVA, R. M. de O.; CARREÑO, L. O. D.; ONO, R. K.; BERTIPAGLIA, T. S.; FARAH, M. M.; MILLEN, D. D.; DA FONSECA, R. Genetic diversity of Brazilian Brahman cattle by pedigree analysis. **Pesquisa Agropecuária Brasileira**, v. 53, n. 1, p. 74–79, 2018.
- COLLEAU, J.-J. An indirect approach to the extensive calculation of relationship coefficients. **Genetics Selection Evolution**, v. 34, n. 4, p. 409–421, jul. 2002. Disponível em: <<http://www.edpsciences.org/10.1051/gse:2002015>>.
- FALCONER, D. S.; MACKAY, T. F. C. **Introduction to Quantitative Genetics (Fourth Edition)**. [s.l: s.n.]v. 12
- FAO. **Secondary guidelines for development of national farm animal genetic resources management plans: management of small populations at risk**. [s.l: s.n.]
- FERRAZ, J. B. S.; ELER, J. P.; GOLDEN, B. L. A formação do Composto Montana Tropical. **Revista Brasileira de Reprodução Animal**, v. 23, n. 2, p. 115–117, 1999a.
- FERRAZ, J. B. S.; ELER, J. P.; GOLDEN, B. L. Análise Genética do Composto Montana Tropical. **Revista Brasileira de Reprodução Animal**, v. 23, n. 2, p. 111–113, 1999b.
- FRANKHAM, R. Conservation Genetics. **Annual Review of Genetics**, v. 29, n. 1, p. 305–327, dez. 1995. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.ge.29.120195.001513>>.
- GRIGOLETTO, L.; BRITO, L. F.; MATTOS, E. C.; ELER, J. P.; BUSSIMAN, F. O.; SILVA, B. da C. A.; DA SILVA, R. P.; CARVALHO, F. E.; BERTON, M. P.; BALDI, F.; FERRAZ, J. B. S. Genome-wide associations and detection of candidate genes for direct and maternal genetic effects influencing growth traits in the Montana Tropical® Composite population. **Livestock Science**, v. 229, p. 64–76, 1 nov. 2019.

GROENEVELD, E.; WESTHUIZEN, B. D.; MAIWASHE, A.; VOORDEWIND, F.; FERRAZ, J. B. POPREP: a generic report for population management.

Genetics and molecular research : GMR, v. 8, n. 3, p. 1158–1178, 2009.

GUTIÉRREZ, J. P.; ALTARRIBA, J.; DÍAZ, C.; QUINTANILLA, R.; CAÑÓN, J.; PIEDRAFITA, J. Pedigree analysis of eight Spanish beef cattle breeds.

Genetics Selection Evolution, v. 35, n. 1, p. 43–63, jan. 2003. Disponível em: <<http://www.gse-journal.org/10.1051/gse:2002035>>.

GUTIÉRREZ, J. P.; CERVANTES, I.; GOYACHE, F. Improving the estimation of realized effective population sizes in farm animals. **Journal of Animal Breeding and Genetics**, v. 126, n. 4, p. 327–332, 2009.

LACY, R. C. Analysis of founder representation in pedigrees: Founder equivalents and founder genome equivalents. **Zoo Biology**, v. 8, n. 2, p. 111–123, 1989. Disponível em: <<http://doi.wiley.com/10.1002/zoo.1430080203>>.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. BLUPF90 and related programs(BGF90). in **Proceedings of the 7th World Congress on Genetics Applied to Livestock Production.**, p. 21–22, 2002.

PEDROSA, V. B.; ELER, J. P.; FERRAZ, J. B. S.; GROENEVELD, E. Use of weaning management group as a random effect for a more robust estimation of genetic parameters for post-weaning traits in Nellore cattle. **Genetics and Molecular Research**, v. 13, n. 3, p. 7013–7021, 2014.

PÉREZ-ENCISO, M. Use of the uncertain relationship matrix to compute effective population size. **Journal of Animal Breeding and Genetics**, v. 112, p. 327–332, 1995.

PERIPOLLI, E.; STAFUZZA, N. B.; AMORIM, S. T.; LEMOS, M. V. A.; GRIGOLETTO, L.; KLUSKA, S.; FERRAZ, J. B. S.; ELER, J. P.; MATTOS, E. C.; BALDI, F. Genome-wide scan for runs of homozygosity in the composite Montana Tropical ® beef cattle. **Journal of Animal Breeding and Genetics**, p. jbg.12428, 8 ago. 2019. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/jbg.12428>>.

PICCOLI, M. L.; BRACCINI NETO, J.; BRITO, F. V.; CAMPOS, L. T.; BÉRTOLI, C. D.; CAMPOS, G. S.; COBUCCI, J. A.; MCMANUS, C. M.; BARCELLOS, J. O. J.; GAMA, L. T. Origins and genetic diversity of British cattle breeds in Brazil assessed by pedigree analyses. **Journal of Animal Science**, v. 92, n. 5, p. 1920–1930, 2014.

PIENAAR, L.; NESER, F. W. C.; GROBLER, J. P.; SCHOLTZ, M. M.; MACNEIL, M. D. Pedigree analysis of the Afrikaner cattle breed. **Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales**, v. 57, p. 51–56, 2015.

R TEAM CORE. **A language and environment for statistical computing.R Foundation for Statistical Computing, Vienna, Austria.** Vienna, Austria, 2019. . Disponível em: <www.R-project.org/>.

SANTANA JR., M. L.; ELER, J. P.; CARDOSO, F. F.; ALBUQUERQUE, L. G.; BALIEIRO, J. C. C.; PEREIRA, R. J.; FERRAZ, J. B. S. Genotype by environment interaction for post-weaning weight gain, scrotal circumference, and muscling score of composite beef cattle in different regions of Brazil. **Genetics and Molecular Research**, v. 13, n. 2, p. 3048–3059, 2014. Disponível em: <<http://www.funpecrp.com.br/gmr/year2014/vol13-2/pdf/gmr2780.pdf>>.

SANTANA, M. L.; ELER, J. P.; BIGNARDI, A. B.; FERRAZ, J. B. S. Genetic associations among average annual productivity, growth traits, and stayability: A parallel between Nelore and composite beef cattle1. **Journal of Animal Science**, v. 91, n. 6, p. 2566–2574, 1 jun. 2013. Disponível em: <<https://academic.oup.com/jas/article/91/6/2566/4717100>>.

SANTANA, M. L.; ELER, J. P.; CARDOSO, F. F.; ALBUQUERQUE, L. G.; BIGNARDI, A. B.; FERRAZ, J. B. S. Genotype by environment interaction for birth and weaning weights of composite beef cattle in different regions of Brazil. **Livestock Science**, v. 149, n. 3, p. 242–249, nov. 2012a. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1871141312002934>>.

SANTANA, M. L.; ELER, J. P.; FERRAZ, J. B. S. Alternative contemporary group structure to maximize the use of field records: Application to growth traits of composite beef cattle. **Livestock Science**, v. 157, n. 1, p. 20–27, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.livsci.2013.06.034>>.

SANTANA, M. L.; OLIVEIRA, P. S.; ELER, J. P.; GUTIÉRREZ, J. P.; FERRAZ, J. B. S. Pedigree analysis and inbreeding depression on growth traits in Brazilian Marchigiana and Bonsmara breeds1. **Journal of Animal Science**, v. 90, n. 1, p. 99–108, 1 jan. 2012b. Disponível em: <<https://academic.oup.com/jas/article/90/1/99/4764689>>.

SARGOLZAEI M., IWAISAKI H., COLLEAU, J. J. CFC : a tool for monitoring genetic diversity . 8th World Congress on Genetics Applied To Livestock. In: 8th World Congress on Genetics Applied to Livestock Production, , Belo Horizonte, MG, Brasil, October, Belo Horizonte - MG. **Anais...** Belo Horizonte - MG: 2006.

UTRERA, Á. R.; MURILLO, V. E. V.; BERMÚDEZ, M. M.; VELÁZQUEZ, G. M.; PONCE, S. I. R. Genetic diversity assessment of the Mexican Simmental population through pedigree analysis. **Revista Brasileira de Zootecnia**, v. 47, 2018.

WRIGHT, S. Mendelian analysis of the pure breeds of livestock IV. The British dairy Shorthorns. **J.Hered.**, v. 17, p. 397–401, 1923.

4. DETECÇÃO E DIAGNÓSTICO DA MULTICOLINEARIDADE EM UMA POPULAÇÃO DE BOVINOS COMPOSTOS MULTIRRACIAIS

Abreu Silva, B.C.¹; Bussiman, F.O.²; Grigoletto, L.¹; Valderrama, A.S.A¹; Eguti, F.C.¹; Mattos, E.C.¹; Ferraz, J.B.S.¹; Eler, J.P.¹

¹ Departamento de Medicina Veterinária; Faculdade de Zootecnia e Engenharia de Alimentos FZEA/USP – Universidade de São Paulo. Av. Duque de Caxias Norte, 225. 13635-900. Pirassununga/SP.

² Departamento de Nutrição e Produção Animal; Faculdade de Medicina Veterinária e Zootecnia FMVZ/USP – Universidade de São Paulo. Av. Duque de Caxias Norte, 225. 13635-900. Pirassununga/SP.

Resumo

O início do programa de formação e melhoramento do Composto Montana[®] ocorreu em meados dos anos 90 com a importação de material genético de diversas raças presentes em todo o mundo, principalmente da África, Austrália, América Central, Caribe e Estados Unidos. O objetivo deste estudo é avaliar a existência da multicolinearidade na população em estudo, bem como quantificar através de diferentes métodos (análise de correlação, decomposição matricial, autovalores) a magnitude dessas relações, propondo alternativas para a correção deste problema na avaliação genética dos bovinos composto Montana[®]. Não existe um método perfeito para a detecção da multicolinearidade, porém, aplicando algumas das diversas metodologias disponíveis, tanto o diagnóstico por meio número de condição da matriz, quanto o método de identificação dos fatores de inflação de variância apontaram que existem relações severas de multicolinearidade para as características ganho de peso (GP) e peso aos 14 meses (P14), já, para a característica peso a desmama (PD) estes métodos apenas alertaram para um forte indicativo de multicolinearidade. Métodos de correção como a regressão de cumeieira, análise de componentes principais e análises de inferência causal podem ser aplicados a fim de minimizar os efeitos da multicolinearidade na avaliação genética desta população

Palavras-chave: Fatores de Inflação de Variância; Correlação; Índice de Condição; Autovalores.

Abstract

The Montana Composite breeding program started the 90's with the import of genetic material from several breeds present throughout the world, mainly from Africa, Australia, Central America, Caribbean and the United States. The aim of this study is to evaluate the existence of multicollinearity in the study population, as well as to quantify the strength of these relationships through different methods (correlation analysis, matrix decomposition, eigenvalues), proposing alternatives for the correction of this problem in the genetic evaluation of Montana® composite beef cattle. There is no perfect method for detecting multicollinearity, however, applying some of the several available methodologies, both the diagnosis by means of the matrix condition number and the method of identification of the variance inflation factors pointed out that there are severe multicollinearity relationships for the post weaning gain (PWG) and yearling weight (YW) traits, for the weaning weight (WW), these methods only warned of a strong indication of multicollinearity. Correction methods such as ridge regression, principal component analysis and causal inference analysis can be applied in order to minimize the effects of multicollinearity on the genetic evaluation of this population.

Keywords : Inflation Variance Factors; Correlation; Condition Index; Eigenvalues.

4.1 Introdução

O Montana, conhecido na mídia rural como Raça Montana, é, na verdade, um bovino composto, desenvolvido com base no cruzamento entre quatro tipos biológicos, cada tipo agrupando algumas raças com características similares. O desenvolvimento do Composto Montana teve como foco principal a produção de animais de alta produtividade de carne, porém com adaptação ao ambiente tropical. Seu desenvolvimento foi também inspirado nos compostos desenvolvidos pelo MARC (*Meat and Animal Research Center*) de Nebraska,

Estados Unidos, tendo como princípio a obtenção e retenção de heterose e complementaridade, a produção de animais com boa uniformidade, com boa qualidade de carne e de carcaça e que fosse simples em termos de manejo. Em se tratando de trópicos, o composto deveria apresentar muita adaptabilidade às condições de ambiente. Os touros deveriam cobrir a campo sem os problemas detectados nos touros de origem europeia (*Bos taurus*). Aqui, então, foi introduzida a principal diferença em relação aos compostos MARC, a utilização de raças europeias adaptadas aos trópicos.

O início do programa de formação e melhoramento do composto se deu em meados dos anos 90 com a importação de material genético de diversas raças presentes em todo o mundo, principalmente da África, Austrália, América Central, Caribe e Estados Unidos. A base de matrizes, no entanto, foram vacas Nelore ou resultante de outros cruzamentos com Nelore. Com muitas raças envolvidas, a inclusão dos efeitos das composições raciais dos produtos, dos pais e das mães, além das heterozigoses, direta e materna, levaria a uma super parametrização do modelo de análise para avaliação genética. Principalmente devido à utilização de vacas com certo grau de cruzamento, foram detectadas raças com pequeno número de observações e raças com grande número de observações. Para atender a um delineamento experimental adequado, principalmente em relação ao número de raças e ao número de animais dentro de cada raça, optou-se pela utilização, na análise genética, do conceito de tipos biológicos em que as raças foram agrupadas de acordo com essa condição. Os animais formadores da população Montana foram agrupados em quatro tipos biológicos. Em um sistema denominado NABC, em que: o tipo biológico N, pertence à classe dos animais zebuínos em sua maioria da raça Nelore que adicionam ao composto características como rusticidade e adaptabilidade aos sistemas de criação em condições adversas, o tipo A é constituído de animais de origem taurina adaptada às condições tropicais, o tipo B corresponde aos animais taurinos britânicos que agregam características como precocidade sexual e de acabamento, assim como qualidade de carne e de carcaça e o tipo C com animais taurinos da região continental da Europa que adicionam ao composto grande desempenho ponderal.

Considerando a utilização do conceito de tipos biológicos, as raças são substituídas nos modelos pelos tipos biológicos e as heterozigoses são

calculadas entre os tipos biológicos, diminuindo muito o número de parâmetros no modelo. Mesmo assim, é possível que se detecte, no composto, o fenômeno da multicolinearidade que ocorre quando as variáveis explanatórias independentes possuem relações lineares muito próximas de 1. Dependendo da magnitude dessas relações, os métodos convencionais de análise, como o método dos quadrados mínimos, podem super ou subestimar os parâmetros devido a um confundimento dos estimadores. Sendo assim, o objetivo deste estudo é avaliar se existe multicolinearidade na população em estudo, bem como, se existir, quantificar, pela aplicação de diferentes métodos, a força dessas relações, propondo alternativas para a correção deste problema na avaliação genética do Composto Montana®.

4.2 Material e Métodos

Por se tratar de um banco comercial e já existente, este estudo foi dispensado da análise ética por utilizar animais oriundos de coleções biológicas formadas anteriormente ao ano de 2008, ano da promulgação da Lei nº 11.794/2008 – lei que estabelece procedimentos para o uso científico de animais.

4.2.1 Banco de dados

Os dados utilizados neste estudo são de bovinos de corte provenientes do programa Montana Composto Tropical®, gerenciado pelo Grupo de Melhoramento Animal e Biotecnologia da Faculdade de Zootecnia e Engenharia de Alimentos da USP. Os bancos de pedigree e fenótipos são compostos de animais nascidos entre os anos de 1994 e 2015, totalizando informações de 562.834 indivíduos, 3.565 touros e 302.044 vacas.

O banco foi editado e foram mantidos apenas os animais com medidas válidas, ou seja, somente animais com informações completas de pedigree para uma correta estimação das heterozigoses direta e materna (Tabela 1). As características analisadas foram peso a desmama (PD, Kg, obtido em torno dos 205 dias), peso ao sobreano (P14, Kg, obtido em torno dos 420 dias) e ganho de peso pós desmama (GP, Kg, obtido pela diferença do peso ao sobreano em relação ao peso à desmama. $GP = (\text{peso ao sobreano} - \text{peso à desmama}) /$

NDEP x 215. NDEP = número de dias entre pesagens e 215 corresponde ao número de dias entre a desmama (205 dias) e o sobreano (420 dias).

Tabela 1 – Estatística descritiva do banco de fenótipos

CARAC.	N	MÉD	MIN	MÁX	DP	NGC	NM	NF
PD	150.773	203,13	70	384	38.59	2.197	78.008	72.765
GP	64.972	78,22	-29.9	477	47.63	962	30.383	34.589
P14	67.902	280,98	102	594	55.97	983	31.505	36.397

N= número de registros, Méd= média, Min= mínimo, Máx= máximo, DP= desvio padrão, NGC= número de grupo de contemporâneos, NM= número de machos, NF= número de fêmeas. Fonte: Própria Autoria

4.2.1.1 Definição da composição racial, heterozigoses e heterozigose materna

Os animais foram classificados de acordo com a sua composição racial, seguindo os tipos biológicos adotado pelo programa, o sistema NABC. Os critérios de classificação estão apresentados na Tabela 2.

Tabela 2 – Classificação dos animais por composição racial para cada característica.

Grupo Genético	NABC	Critério (%)	NPD	NGP	NP14
Outros	XXXX	Qualquer composição que não atenda às condições abaixo	44.709	19.998	20.864
$\frac{3}{4}$	12XXX	60< N <90	2.832	489	546
	X12XX	60< A <90	14.906	8.101	8.386
	XX12X	60< B <90	2.211	957	1.086
	XXX12	60< C <90	176	13	13
Raças Puras	16000	N ≥90	2.297	37	45
	01600	A ≥90	1.059	606	614
	00160	B ≥90	1.536	478	553
	00016	C ≥90	1	0	0
F ₁	8800	40 ≤ N ≤60 e 40 ≤ A ≤60	978	284	296
	8080	40 ≤ N ≤60 e 40 ≤ B ≤60	15.616	3.370	3.500
	8008	40 ≤ N ≤60 e 40 ≤ C ≤60	3.471	542	580
	0880	40 ≤ A ≤60 e 40 ≤ B ≤60	1.746	1.079	1.117
	0808	40 ≤ A ≤60 e 40 ≤ C ≤60	20	6	6
Montana		18,75< N <31,25 e 18,75< A <31,25 e 18,75< B <31,25 e 18,75< C <31,25	8.403	3.877	4.165
	4444				
	4480	18,75< N <31,25 e 18,75< A <31,25 e 43,75< B <56,25	2.068	1.103	1.211
	4804	18,75< N <31,25 e 18,75< A <31,25 e 18,75< C <31,25	10.368	5.175	5.435
	4840	18,75< N <31,25 e 43,75< A <56,25 e 18,75< B <31,25	38.161	18.741	19.363
		N ≤37,5 e 12,5 ≤ A ≤87,5 e B ≤75 e C ≤75 e (N + A) ≤0,25 e (B + C) ≤75	215	116	122
	MMMM				

NPD= número de registros para o peso a desmama, NGP= número de registros para o ganho de peso, NP14= número de registros para o peso ao sobreano. Fonte: Própria Autoria

Para a inclusão nos modelos de análise temos as covariáveis:

A; B e C = que se referem a composição racial do animal (dadas como desvio do tipo biológico N);

Am, Bm e Cm = que são as composições raciais maternas (dadas como desvio do tipo biológico materno N).

A equação para o cálculo das heterozigoses foi dada por:

$$HTZ = \sum_{i \neq j} r_i r_j$$

em que, r_i é a proporção do i-ésimo tipo biológico paterno, r_j é a proporção do j-ésimo tipo biológico materna. Logo em seguida foram classificadas, para posterior inclusão nas análises como:

NA = heterozigose referente as raças do grupo N e do grupo A existente na composição racial do animal;

NB = heterozigose referente as raças do grupo N e do grupo B existente na composição racial do animal;

NC = heterozigose referente as raças do grupo N e do grupo C existente na composição racial do animal;

AB = heterozigose referente as raças do grupo A e do grupo B existente na composição racial do animal;

AC = heterozigose referente as raças do grupo A e do grupo C existente na composição racial do animal;

BC = heterozigose referente as raças do grupo B e do grupo C existente na composição racial do animal;

HTM = a heterozigose materna total, é dada como a heterozigose total das fêmeas quando se tornam mães.

4.2.2 Identificação da multicolinearidade

O diagnóstico da multicolinearidade foi realizado com base na aplicação das metodologias descritas abaixo em uma matriz que **X** que continha as informações agrupadas das composições raciais dos animais (A, B, C), da composição racial materna dos animais (Am, Bm e Cm), heterozigoses (NA, NB, NC, AB, AC, BC) e heterozigose materna total (HTM).

4.2.2.1 Análise de Correlação

Utilizada para medir a associação entre duas variáveis, a correlação é utilizada como um indicativo de multicolinearidade quando as estimativas dos coeficientes se encontram muito próximas de 1 ou -1. Para esse estudo, foram obtidas as estimativas do coeficiente de correlação de Pearson na matriz $\mathbf{X}$. Para uma melhor visualização gráfica dos resultados, utilizou-se o pacote CORRPLOT (WEI; SIMKO, 2017) do software estatístico R (R CORE TEAM, 2018).

4.2.2.2 Número de Condição (NC) e Índice de Condição (IC) da Matriz

Dado como a razão entre o maior e o menor autovalor da matriz de correlação $\mathbf{X}'\mathbf{X}$ (MONTGOMERY & PECK, 1982), o NC é um indicativo de multicolinearidade. Para a classificação da força dessas relações, considerou-se a metodologia descrita por Cruz & Carneiro (2003), em que, se o $NC < 100$, a multicolinearidade é considerada fraca e não constitui um grave problema para as análises; se $100 \leq NC \leq 1000$, a multicolinearidade é considerada moderada a forte e se $NC \geq 1000$, é considerada severa. Outra maneira de quantificar e determinar a força das colinearidades, consiste em identificar os índices de condição, para $10 \leq IC \leq 30$, dizemos que há uma relação de quase dependência, $IC > 30$ são indicativos de colinearidades fortes (DIAS, 2008).

4.2.2.3 Decomposição de variância dos coeficientes de regressão

Este método permite relacionar o condicionamento da matriz de dados, determinado pela decomposição de cada valor singular com a qualidade da análise de regressão. Proporções elevadas de dois ou mais coeficientes pertencentes a valores singulares pequenos, indicam que as quase dependências observadas podem causar problemas. O primeiro passo é a detecção dos autovalores da matriz $\mathbf{X}$, e segundo a metodologia de Belsley et al. (1980), proporções de decomposição de variância acima de 0,50 para os autovalores indicam a relação de colinearidade. Para os cálculos de

determinação dos autovalores e autovetores da matriz de covariáveis foi utilizado o pacote OLSRR (HEBBALI, 2018) do software R (R Development Core Team, 2016).

4.2.2.4 Fatores de Inflação de variância (FIV)

Este método de identificação da multicolinearidade mede o quanto as variações dos coeficientes de regressão estimados estão infladas se comparados com as variáveis preditoras que não são linearmente relacionados. Calculado por: $FIV = \frac{1}{1-R_i^2}$ em que, R_i^2 é o coeficiente de correlação múltipla da regressão linear das variáveis explanatórias (**X**). Quando as variáveis de **X** não são colineares temos um $R_i^2 = 0$ e consequentemente o $FIV=1$, quando $R_i^2 \neq 0$ o $FIV>1$. Para este estudo foram considerados como indicação de colinearidade $FIV > 10$.

4.3 Resultados e Discussão

4.3.1 Correlações

Foram encontradas algumas correlações altas entre as variáveis da matriz **X**, contudo os padrões encontrados se diferiram um pouco entre características, conforme demonstrado na Figura 1. Para o PD se destacaram duas correlações altas ($>0,70$), entre as variáveis AB x Bm (0,78) e entre AC x Cm (0,88). Provavelmente, essa associação pode existir pois os componentes maternos (Bm e Cm) fazem parte da formação das heterozigoses do animal (AB e AC).

Figura 1- Plotagem da matriz de correlação entre as variáveis de X.

Fonte: Própria autoria.

Tanto para a característica GP, quanto para a característica P14, os padrões observados graficamente demonstram que variáveis apresentam altas correlações, a maior delas foi a que envolve as variáveis AC x Cm (0,90), seguidos por AB x Bm (0,80), C x Cm (0,77), C x AC (0,73) e NB x A (-0,70). Alguns autores concluem que só se deve utilizar esse diagnóstico de multicolinearidade por meio da análise de correlações se os valores encontrados demonstrarem uma relação linear perfeita, ou seja, se forem 1 ou -1 (GUAJARATI; POTTER, 2011; NETER; WASSERMAN, 1974).

Apesar da utilidade das estimativas do coeficiente de correlação para a compreensão de um caráter complexo, essas estimativas não determinam, entretanto, a importância relativa das influências diretas e indiretas dos outros caracteres na produção, pois a correlação entre duas características mede a associação entre ambas, mas não determina a relação de causa e efeito entre elas.

4.3.2 NC, IC e decomposição da variância da matriz X

O diagnóstico para colinearidade pelos números de condição encontrados foram 809,13; 6.317,95 e 5.670,47 para as matrizes de PD, GP e P14, caracterizando uma multicolinearidade severa para todas as características.

As Tabelas 3, 4 e 5 apresentam os resultados referentes a decomposição das matrizes das covariáveis para as características PD, GP e P14 respectivamente. Para a característica PD (Tabela 3) os autovalores 0,0085;

0,0126; 0,0144; 0,0294; 0,0574; apresentaram os respectivos IC associados: 28,4453; 23,3867; 21,8272; 15,2934; 10,9464, indicando relações de quase dependências, ou seja, estes resultados apontam um indicativo de colinearidade fraca. Malau-Aduli et al. (2004) sugeriram avaliar os pequenos autovalores, justamente por serem capazes de identificar as altas proporções de variância e assim proferir um diagnóstico mais preciso para a multicolinearidade. Dias et al. (2012), em estudos com animais Montana, encontrou dois autovalores com dependências fortes para a mesma característica em questão.

Para a característica GP (Tabela 4) os autovalores 0,0011; 0,0066; 0,0106; 0,0294 e 0,0503; apresentaram os respectivos IC associados: 79,4855; 32,9161; 26,0289; 15,6217 e 11,9603, os dois primeiros autovalores apresentaram um IC acima de 30 caracterizando um indicativo forte de colinearidade, já os três últimos indicaram apenas relações de quase dependências. A partir da decomposição proporcional da variância, se considerarmos o limiar de 0,5, na linha associada ao IC de 79,4855, as covariáveis A (0,5961), NA (0,73), NB (0,79), corroborando mais uma vez com um forte indicativo de multicolinearidade. Se considerarmos a linha do autovalor 32,9161 verificamos relações de colinearidade forte para as variáveis C (0,8323), NC (0,6091), AC (0,6020) e BC (0,5823). Em uma população africana de bovinos *White Fulani*, Yakubu et al. (2010), encontraram índices de condição que indicaram fortes dependências para variáveis de crescimento corporal, os autores evidenciaram a multicolinearidade através desse método.

Padrão semelhante foi observado para a característica P14 (Tabela 5), em que, os autovalores 0,0013; 0,0067; 0,0109; 0,0295 e 0,0509; apresentaram os respectivos IC associados: 75,3025; 32,8515; 25,7231; 15,5936 e 11,8795. A partir da decomposição proporcional da variância, se considerarmos a linha associada ao IC de 75,3025, temos forte colinearidade entre as covariáveis A (0,58), NA (0,7090), NB (0,7830). Para a linha do autovalor 32,8515 verificamos relações de colinearidade forte para as variáveis C (0,8493), NC (0,6466), AC (0,6407) e BC (0,6211).

Tabela 3 – Decomposição proporcional da variância PD das estimativas dos coeficientes de regressão

	Auto Valor	IC	Int	A	B	C	Am	Bm	Cm	NA	NB	NC	AB	AC	BC	HTm
INT	6,8781	1,0000	0,0003	0,0004	0,0004	0,0001	0,0006	0,0004	0,0005	0,0010	0,0002	0,0000	0,0011	0,0003	0,0008	0,0020
A	2,2451	1,7503	0,0000	0,0002	0,0006	0,0014	0,0001	0,0018	0,0053	0,0004	0,0000	0,0011	0,0055	0,0032	0,0039	0,0003
B	1,5226	2,1254	0,0002	0,0010	0,0047	0,0001	0,0013	0,0001	0,0009	0,0024	0,0190	0,0021	0,0004	0,0019	0,0052	0,0015
C	1,1336	2,4632	0,0000	0,0000	0,0010	0,0010	0,0013	0,0002	0,0040	0,0000	0,0064	0,0207	0,0011	0,0021	0,0052	0,0000
Am	0,9060	2,7553	0,0002	0,0010	0,0000	0,0001	0,0597	0,0040	0,0006	0,0016	0,0080	0,0000	0,0053	0,0000	0,0117	0,0001
Bm	0,7492	3,0300	0,0004	0,0000	0,0000	0,0000	0,0196	0,0008	0,0004	0,0072	0,0115	0,0009	0,0000	0,0037	0,1023	0,0002
Cm	0,2431	5,3195	0,0003	0,0010	0,0169	0,0006	0,0068	0,0287	0,0025	0,0798	0,0124	0,0004	0,0012	0,0254	0,0677	0,0205
NA	0,1060	8,0556	0,0041	0,0148	0,0024	0,0001	0,0043	0,0060	0,0193	0,0931	0,0062	0,0001	0,0472	0,0044	0,0250	0,6270
NB	0,0940	8,5524	0,0006	0,0129	0,0260	0,0002	0,0000	0,0204	0,0143	0,0414	0,0344	0,0000	0,5364	0,0000	0,0124	0,1782
NC	0,0574	10,9464	0,0052	0,0023	0,0393	0,0124	0,0009	0,0037	0,5191	0,0001	0,0545	0,0082	0,0103	0,2053	0,0605	0,0441
AB	0,0294	15,2934	0,0334	0,2078	0,2663	0,0032	0,1857	0,1712	0,0058	0,2939	0,0844	0,0000	0,3080	0,0173	0,0032	0,0175
AC	0,0144	21,8272	0,2689	0,5712	0,4288	0,0545	0,2435	0,1455	0,0421	0,0464	0,0200	0,1516	0,0516	0,0809	0,0986	0,0873
BC	0,0126	23,3867	0,6709	0,0039	0,0418	0,0986	0,1980	0,3831	0,1817	0,3276	0,5819	0,0082	0,0208	0,1533	0,1024	0,0134
HTm	0,0085	28,4453	0,0154	0,1835	0,1718	0,8276	0,2782	0,2341	0,2035	0,1053	0,1610	0,8068	0,0111	0,5020	0,5012	0,0079

IC= índice de condição, Int= Intercepto, A=Composição racial do animal do tipo biológico adaptado, B= Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

Tabela 4 - Decomposição proporcional da variância GP das estimativas dos coeficientes de regressão

	Auto Valor	IC	Int	A	B	C	Am	Bm	Cm	NA	NB	NC	AB	AC	BC	HTm
INT	7,1913	1,0000	0,0000	0,0001	0,0003	0,0001	0,0003	0,0002	0,0003	0,0003	0,0001	0,0001	0,0009	0,0002	0,0005	0,0017
A	2,3098	1,7645	0,0000	0,0000	0,0004	0,0010	0,0000	0,0010	0,0033	0,0001	0,0000	0,0007	0,0047	0,0019	0,0026	0,0003
B	1,3947	2,2707	0,0000	0,0004	0,0024	0,0000	0,0008	0,0001	0,0004	0,0008	0,0092	0,0025	0,0002	0,0011	0,0094	0,0012
C	1,0325	2,6391	0,0000	0,0000	0,0004	0,0003	0,0040	0,0000	0,0021	0,0001	0,0023	0,0301	0,0000	0,0010	0,0018	0,0000
Am	0,8915	2,8401	0,0000	0,0004	0,0000	0,0002	0,0305	0,0018	0,0002	0,0007	0,0050	0,0022	0,0046	0,0000	0,0018	0,0000
Bm	0,6766	3,2601	0,0000	0,0000	0,0000	0,0000	0,0059	0,0005	0,0001	0,0016	0,0103	0,0033	0,0001	0,0014	0,0797	0,0000
Cm	0,2336	5,5487	0,0000	0,0003	0,0124	0,0007	0,0038	0,0127	0,0018	0,0224	0,0045	0,0003	0,0023	0,0135	0,0535	0,0347
NA	0,0914	8,8719	0,0003	0,0101	0,0026	0,0002	0,0006	0,0005	0,0056	0,0177	0,0007	0,0000	0,0047	0,0012	0,0259	0,7807
NB	0,0805	9,4534	0,0002	0,0010	0,0269	0,0000	0,0003	0,0134	0,0295	0,0277	0,0165	0,0003	0,5942	0,0016	0,0010	0,0020
NC	0,0503	11,9603	0,0003	0,0012	0,0489	0,0186	0,0001	0,0086	0,3330	0,0011	0,0258	0,0144	0,0187	0,1018	0,0208	0,0509
AB	0,0294	15,6517	0,0006	0,0791	0,1257	0,0048	0,0847	0,1255	0,0236	0,0752	0,0280	0,0017	0,3100	0,0140	0,0009	0,0339
AC	0,0106	26,0289	0,0000	0,1953	0,2678	0,1063	0,3253	0,3160	0,2100	0,0860	0,0700	0,1071	0,0012	0,2519	0,2174	0,0608
BC	0,0066	32,9161	0,0004	0,1159	0,1073	0,8223	0,2166	0,1445	0,1197	0,0362	0,0339	0,6091	0,0143	0,6020	0,5823	0,0021
HTm	0,0011	79,4855	0,9981	0,5961	0,4048	0,0454	0,3270	0,3752	0,2707	0,7300	0,7935	0,2282	0,0440	0,0083	0,0024	0,0316

IC= índice de condição, Int= Intercepto, A=Composição racial do animal do tipo biológico adaptado, B= Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

Tabela 5 – Decomposição proporcional da variância P14 das estimativas dos coeficientes de regressão

	Auto Valor	IC	Int	A	B	C	Am	Bm	Cm	NA	NB	NC	AB	AC	BC	HTm
INT	7,1824	1,0000	0,0000	0,0002	0,0003	0,0001	0,0003	0,0002	0,0003	0,0003	0,0001	0,0001	0,0009	0,0002	0,0005	0,0017
A	2,3065	1,7647	0,0000	0,0001	0,0004	0,0010	0,0000	0,0010	0,0034	0,0001	0,0000	0,0007	0,0049	0,0019	0,0025	0,0003
B	1,3954	2,2687	0,0000	0,0005	0,0026	0,0000	0,0009	0,0001	0,0004	0,0009	0,0100	0,0023	0,0003	0,0011	0,0092	0,0013
C	1,0359	2,6331	0,0000	0,0000	0,0004	0,0003	0,0044	0,0000	0,0021	0,0001	0,0022	0,0309	0,0000	0,0010	0,0018	0,0000
Am	0,8907	2,8397	0,0000	0,0004	0,0000	0,0002	0,0325	0,0019	0,0002	0,0007	0,0058	0,0023	0,0048	0,0000	0,0016	0,0000
Bm	0,6758	3,2600	0,0000	0,0000	0,0000	0,0000	0,0066	0,0006	0,0001	0,0018	0,0117	0,0032	0,0001	0,0014	0,0763	0,0000
Cm	0,2378	5,4963	0,0000	0,0003	0,0133	0,0007	0,0041	0,0132	0,0018	0,0235	0,0051	0,0003	0,0014	0,0134	0,0528	0,0341
NA	0,0919	8,8426	0,0003	0,0108	0,0031	0,0002	0,0007	0,0004	0,0061	0,0184	0,0011	0,0000	0,0075	0,0013	0,0255	0,7796
NB	0,0844	9,2234	0,0002	0,0008	0,0271	0,0000	0,0003	0,0120	0,0257	0,0324	0,0176	0,0002	0,5890	0,0008	0,0007	0,0006
NC	0,0509	11,8795	0,0003	0,0013	0,0483	0,0184	0,0002	0,0107	0,3462	0,0012	0,0262	0,0144	0,0106	0,0997	0,0196	0,0566
AB	0,0295	15,5936	0,0007	0,0859	0,1300	0,0049	0,0906	0,1291	0,0261	0,0817	0,0289	0,0018	0,3217	0,0150	0,0009	0,0309
AC	0,0109	25,7231	0,0000	0,2155	0,2906	0,0828	0,3584	0,3439	0,2278	0,0970	0,0774	0,0885	0,0014	0,2154	0,1850	0,0628
BC	0,0067	32,8515	0,0005	0,1043	0,0936	0,8493	0,1968	0,1281	0,1042	0,0328	0,0309	0,6466	0,0134	0,6407	0,6211	0,0010
HTm	0,0013	75,3025	0,9978	0,5800	0,3903	0,0420	0,3042	0,3587	0,2557	0,7090	0,7830	0,2089	0,0440	0,0081	0,0024	0,0311

IC= índice de condição, Int= Intercepto, A=Composição racial do animal do tipo biológico adaptado, B= Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

Analisando os FIV (Tabela 6), para a característica PD as variáveis que contribuem para a multicolinearidade, se considerarmos que há multicolinearidade caso os valores de FIV sejam maiores que 10, são A, C, Bm, Cm, NB, NC e AC. O mesmo padrão foi encontrado por Dias (2008), exceto para a variável B que também foi encontrada em seu estudo. Para as características GP e P14, todas as variáveis apresentaram alto FIV, exceto para AB, BC e HTM. As estimativas dos coeficientes de regressão encontrados apresentam um padrão de inflação bastante significativo. As relações entre as composições maternas e diretas e as heterozigoses já foram alvos vários de estudos, Rodríguez-Almeida; Van Vleck e Gregory (1997) destacaram a importância da relação entre as composições raciais maternas e diretas em animais multirraciais, os autores ainda apontam que para que se possa aplicar modelos mais simplificados na avaliação é necessário que se tenha informações completas, além de uma variedade nos cruzamentos, se não a estimativa destes efeitos pode ser prejudicada. Esta relação de confundimento e não possibilidade de estimação destes efeitos pode ser explicadas em grande parte pela forte presença de multicolinearidade para estas variáveis.

Tabela 6 – Fatores de inflação de variância (FIV).

COV	PD	GP	P14
A	11,07	19,31	18,22
B	9,05	14,60	13,99
C	40,68	52,87	52,96
Am	7,86	14,41	13,44
Bm	13,66	22,16	21,12
Cm	11,94	19,12	18,36
NA	5,59	15,40	14,08
NB	10,13	23,84	21,67
NC	22,87	19,64	19,34
AB	5,24	5,05	4,92
AC	16,82	28,22	28,37
BC	5,67	7,87	8,10
HTM	2,33	1,70	1,70

COV= covariável, PD= peso à desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou sobreano, A=Composição racial do animal do tipo biológico adaptado, B=

Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

Petrini et al. (2012) encontraram para P14, FIVs maiores que 10 para as variáveis A, C, NC e AC, os autores sugeriram uma análise mais detalhada para possível exclusão de variáveis a fim de sanar o problema da multicolinearidade naquela população. Avaliando a presença ou não de multicolinearidade em uma população de bovinos cruzados, com diferentes proporções de cruzamentos de animais Angus e Nelore, Bertoli et al. (2015), só encontraram indícios de multicolinearidade na variável perdas epistáticas para a característica perímetro escrotal. Malau-Aduli et al. (2004), encontraram severas relações de multicolinearidade entre características de escore e peso corporal em bovinos da raça *Japanese Black Cattle*, os autores associaram o grau de multicolinearidade como severo para FIVs entre 21,89 a 46,94 para as variáveis em estudo.

Apenas as variáveis AB, BC e HTM aparentemente não contribuíram para a multicolinearidade nas características. As técnicas do NC e IC informam o grau de multicolinearidade na matriz de correlação $X'X$, já a técnica do FIV e do exame dos autovalores informa o quanto cada variável explicativa contribui para a multicolinearidade. Apesar dos métodos de diagnóstico se diferirem em suas particularidades, para as características envolvidas neste estudo, houve um padrão de semelhança que separou o diagnóstico de multicolinearidade na característica PD para com as demais características (GP e P14), todas as análises apontaram que o PD aparentemente é menos influenciado pela multicolinearidade do que as variáveis P14 e GP. A análise de correlação e a análise dos FIV para o PD apontaram uma similaridade na relação que ocorre entre as variáveis AC, NC, NB, Cm, o que não indica a certeza da existência de multicolinearidade, mas também não descarta a mesma. Para o GP e o P14 o padrão entre todos os métodos foram bastante semelhantes, evidenciando que o diagnóstico para a multicolinearidade para estas características foi suficientemente grande para provar que existe este fenômeno e merece atenção

especial nas análises de estimação dos componentes de variância pelo método dos quadrados mínimos e na predição dos valores genéticos para estas características.

4.4 Considerações Finais

Apesar da interpretação dos coeficientes de correlação exigirem um elevado grau de atenção, devido à complexidade de compreensão das relações observadas, foram encontrados indícios muito fortes de multicolinearidade envolvendo algumas variáveis. Um estudo de causa e efeito pode ser indicado para identificar a profundidade da relação destas variáveis, com isso, pode-se eliminar algumas das covariáveis do modelo de avaliação genética (reduzindo então a parametrização do mesmo), diminuindo e/ou eliminando os efeitos da multicolinearidade.

Não existe um método perfeito para a detecção da multicolinearidade, porém, aplicando algumas das diversas metodologias disponíveis, tanto o diagnóstico por meio número de condição da matriz, quanto o método de identificação dos fatores de inflação de variância apontaram que existem relações severas de multicolinearidade para as características GP e P14, já, para a característica PD estes métodos apenas alertaram para um forte indicativo de multicolinearidade. Métodos de correção como a regressão de cumeieira, análise de componentes principais e análises de inferência causal podem ser aplicados a fim de minimizar os efeitos da multicolinearidade na avaliação genética desta população.

Referências

- BELSLEY, D. A.; KUH, E.; WELSCH, R. E. Identifying influential data and sources of collinearity. **Regression Diagnostics**, 1980.
- BERTOLI, C. D.; BRACCINI, J.; MCMANUS, C.; COBUCI, J. A.; KERN, E. L.; PICCOLI, M. L.; SCHENKEL, F.; ROSO, V. Modeling breed additive and non-additive genetic effects using a Angus x Nellore crossbred population. **Livestock Science**, v. 176, n. 2015, p. 1–13, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.livsci.2015.03.020>>.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. [s.l: s.n.]
- DIAS, R. A. P. **Diagnóstico e redução da influência da multicolinearidade na estimação de efeitos genéticos aditivos e não-aditivos em uma população de bovinos compostos (Bos taurus x Bos indicus)**. 2008. scola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2008.
- DIAS, R. A. P.; PETRINI, J.; FERRAZ, J. B. S.; ELER, J. P.; BUENO, R. S.; DA COSTA, A. L. L.; MOURÃO, G. B. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. **Livestock Science**, v. 142, n. 1–3, p. 188–194, dez. 2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1871141311002836>>.
- GUJARATI, D. N.; PORTER., D. C. **Econometria Básica-5**. [s.l.] Amgh Editora, 2011.
- HEBBALI, A. **olsrr: Tools for Building OLS Regression Models**, 2018. . Disponível em: <<https://cran.r-project.org/package=olsrr>>.
- MALAU-ADULI, A. E. O.; AZIZ, M. A.; KOJIMA, T.; NIIBAYASHI, T.; OSHIMA, K.; KOMATSU, M. Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle. **Journal of Animal and Veterinary Advances**, v. 3, n. 12, p. 856–863, 2004.
- MONTGOMERY, DOUGLAS C. PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis, 5th Edition**. [s.l.] John Wiley & Sons, 2012.
- NETER, J.; WASSERMAN, W.; KUTNER, M. H. **Applied linear statistical models. Richard D. Irwin. Inc., Home ed.** [s.l: s.n.]
- PETRINI, J.; ANTONIO, R.; DIAS, P.; FERNANDA, S.; PERTILE, N.; ELER, J. P.; BENTO, J.; FERRAZ, S.; MOURÃO, G. B. Degree of multicollinearity and variables involved in linear dependence in additive - dominant models. n. 1, p. 1743–1750, 2012.
- R TEAM CORE. **A language and environment for statistical computing.R Foundation for Statistical Computing, Vienna, Austria**.Vienna, Austria, 2019. . Disponível em: <www.R-project.org/>.

RODRÍGUEZ-ALMEIDA, F. A.; VAN VLECK, L. D.; GREGORY, K. E. Estimation of direct and maternal breed effects for prediction of expected progeny differences for birth and weaning weights in three multibreed populations. **Journal of Animal Science**, v. 75, n. 5, p. 1203, 1997. Disponível em: <<https://academic.oup.com/jas/article/75/5/1203-1212/4624999>>.

WEI, T.; SIMKO, V. **R package “corrplot”: Visualization of a Correlation Matrix (Version 0.84)**, 2017. . Disponível em: <<https://github.com/taiyun/corrplot>>.

YAKUBU, A. Fixing multicollinearity instability in the prediction of body weight from morphometric traits of White Fulani cows. **Journal of Central European Agriculture**, v. 11, n. 4, p. 487–492, 2010.

5. MÉTODOS DE TRATAMENTO PARA A MULTICOLINEARIDADE EM UMA AVALIAÇÃO GENÉTICA DE BOVINOS COMPOSTOS MULTIRRACIAIS

Abreu Silva, B.C.¹; Bussiman, F.O.²; Grigoletto, L.¹; Valderrama, A.S.A¹; Eguti, F.C.¹; Mattos, E.C.¹; Ferraz, J.B.S.¹; Eler, J.P.¹

¹ *Departamento de Medicina Veterinária; Faculdade de Zootecnia e Engenharia de Alimentos FZEA/USP – Universidade de São Paulo. Av. Duque de Caxias Norte, 225. 13635-900. Pirassununga/SP.*

² *Departamento de Nutrição e Produção Animal; Faculdade de Medicina Veterinária e Zootecnia FMVZ/USP – Universidade de São Paulo. Av. Duque de Caxias Norte, 225. 13635-900. Pirassununga/SP.*

Resumo

O Montana, conhecido na mídia rural como Raça Montana, é, na verdade, um bovino composto, desenvolvido com base no cruzamento entre quatro tipos biológicos, cada tipo agrupando algumas raças com características similares. Para atender a um delineamento experimental adequado, principalmente em relação ao número de raças e ao número de animais dentro de cada raça, optou-se pela utilização, na análise genética, do conceito de tipos biológicos em que as raças foram agrupadas de acordo com essa condição. A primeira etapa de um processo de avaliação genética consiste na estimação dos componentes de (co)variância e dos parâmetros genéticos da população e, nos programas de melhoramento genético de animais cruzados e compostos, deve-se levar em conta a multicolinearidade, um fator que afeta diretamente a precisão das estimativas desses parâmetros genéticos. A multicolinearidade é um fator relacionado com a presença de fortes relações lineares entre as variáveis explanatórias e que pode levar a uma instabilidade nos coeficientes de regressão, aumentando o erro padrão associado aos componentes, sobretudo quando a metodologia de análise é a dos quadrados mínimos. O objetivo deste estudo é obter e comparar as estimativas dos efeitos genéticos aditivos direto e materno, assim como dos efeitos fixos, pelos métodos sem correção (SC), máxima verossimilhança restrita (REML), quadrados mínimos (QM), regressão

de cumeeira (RC), análise de fatores (AF) e componentes principais (PCA). As correções para a multicolinearidade efetuadas foram eficientes para ajustar o β das covariáveis, para todos os métodos. Por mais que os métodos apresentaram diferenças nas análises de comparação de modelos, como o ajuste do β das covariáveis não interfere na predição dos valores genéticos, recomenda-se a aplicação da correção para a estimação dos componentes de variância e dos parâmetros genéticos, cabe ao pesquisador então um maior detalhamento do seu banco de dados para a escolha do modelo parcimonioso mais adequado.

Palavras- Chave: Colinearidade; Regressão de Cumeeira; Análise de Fatores; Análise de Componentes Principais.

Abstract

Montana, known in the rural media as the Montana Breed, is actually a composite beef cattle, developed based on the cross between four biological types, each type grouping some breeds with similar traits. In order to attend to an adequate experimental design, mainly in relation to the number of breeds and the number of animals within each breed, it was decided to use, in genetic analysis, the concept of biological types in which the breeds were grouped according to this condition. The first stage of a genetic evaluation process consists of estimating the components of (co) variance and the genetic parameters of the population and, in the breeding programs of crossbred and composite animals, multicollinearity must be taken into account, a factor that directly affects the accuracy of estimates of these genetic parameters. Multicollinearity is a factor related to the presence of strong linear relationships between explanatory variables and that can lead to instability in the regression coefficients, increasing the standard error associated with the components, especially when the analysis methodology is that of the least squares. The objective of this study is to obtain and compare the estimates of direct and maternal additive genetic effects, as well as fixed effects, by methods without correction (WC), restricted maximum likelihood (REML), ordinary least squares (OLS), ridge regression (RR), factor analysis (FA) and principal components regression (ACP). The corrections for

multicollinearity performed were efficient to adjust the β of the covariates, for all methods. As much as the methods showed differences in the model comparison analyzes, as the adjustment of the β of the covariates does not interfere in the prediction of genetic values, it is recommended to apply the correction to estimate the variance components and genetic parameters, it is up to the researcher then gave more details to his database to choose the most appropriate parsimonious model.

Key words: Collinearity; Ridge regression; Factor analysis; Principal component regression.

5.1 Introdução

Os cruzamentos têm se constituído em uma ferramenta de grande importância para obtenção de melhores índices de produtividade na pecuária de corte. Dentre as estratégias de cruzamento, a formação e utilização de bovinos composto é uma alternativa viável com uma série de fatores favoráveis. Os bovinos compostos são formados pelo cruzamento de duas ou mais raças de forma a obter animais que retém muita heterose e possuem boa complementaridade e uniformidade. O ponto de partida para um programa de seleção é a avaliação genética de todos os animais. Os resultados da avaliação genética são fundamentais para a tomada de decisão de acasalamento. A primeira etapa de um processo de avaliação genética consiste na estimação dos componentes de (co)variância e dos parâmetros genéticos da população e, nos programas de melhoramento genético de animais cruzados e compostos, deve-se levar em conta a multicolinearidade, um fator que afeta diretamente a precisão das estimativas desses parâmetros genéticos.

A multicolinearidade é um fator relacionado com a presença de fortes relações lineares entre as variáveis explanatórias e que pode levar a uma instabilidade nos coeficientes de regressão, aumentando o erro padrão associado aos componentes, sobretudo quando a metodologia de análise é a dos quadrados mínimos. Em bovinos de corte, existem poucos estudos sobre a

multicolinearidade causada pela alta parametrização dos modelos para estimação dos efeitos genéticos em populações cruzadas, sobretudo pela inclusão das composições raciais e das heterozigoses (KINGHORN; VERCOE, 1989; SCHOEMAN; AZIZ; JORDAAN, 2002; PIMENTEL, 2004; ROSO et al., 2005b; DIAS et al., 2011; PETRINI et al., 2012; BERTOLI et al., 2019).

Na presença de multicolinearidade, as estimativas dos componentes de (co)variância podem assumir valores absurdos, ou completamente diferentes do encontrado para as mesmas características em populações semelhantes. Os resultados então se tornam muito sensíveis a mudanças nos bancos de dados, ou modificações na estrutura de parametrização dos modelos (BELSLEY; KUH; WELSCH, 1980). Estimar corretamente os parâmetros genéticos da população constitui-se em uma condição fundamental para o processo de predição do mérito genético de cada animal da população. O objetivo deste estudo é obter e comparar as estimativas dos efeitos genéticos aditivos direto e materno, assim como dos efeitos fixos, pelos métodos sem correção (SC), máxima verossimilhança restrita (REML), quadrados mínimos (QM), regressão de cumeira (RC), análise de fatores (AF) e componentes principais (PCA).

5.2 Material e Métodos

5.2.1 Banco de dados

Dados fenotípicos e de pedigree de 562.834 indivíduos, 3.565 touros e 302.044 vacas, nascidos entre 1994 e 2015. O banco foi editado e foram mantidos apenas os animais com medidas válidas, ou seja, somente animais com informações completas de pedigree para uma correta estimação das heterozigoses direta e materna. As características analisadas (Tabela 1) foram peso a desmama (PD, Kg, obtido em torno dos 205 dias), peso ao sobreano (P14, Kg, obtido em torno dos 420 dias) e ganho de peso pós desmama (GP, Kg, obtido pela diferença do peso ao sobreano em relação ao peso à desmama. $GP = (\text{peso ao sobreano} - \text{peso à desmama}) / \text{NDEP} \times 215$. NDEP = número de dias entre pesagens e 215 corresponde ao número de dias entre a desmama (205 dias) e o sobreano (420 dias).

Tabela 1 – Estatística descritiva do banco de fenótipos

CARAC.	N	MÉD	MIN	MÁX	DP	NGC	NM	NF
PD	150.773	203,13	70	384	38.59	2.197	78.008	72.765
GP	64.972	78,22	-29.9	477	47.63	962	30.383	34.589
P14	67.902	280,98	102	594	55.97	983	31.505	36.397

N= número de registros, Méd= média, Min= mínimo, Máx= máximo, DP= desvio padrão, NGC= número de grupo de contemporâneos, NM= número de machos, NF= número de fêmeas.

Fonte: Própria Autoria

Os dados utilizados neste estudo são de bovinos de corte provenientes do programa Composto Montana[®], gerenciado pelo Grupo de Melhoramento Animal e Biotecnologia da Faculdade de Zootecnia e Engenharia de Alimentos da USP. Por se tratar de um banco comercial e já existente, este estudo foi dispensado da análise ética por utilizar animais oriundos de coleções biológicas formadas anteriormente ao ano de 2008, ano da promulgação da Lei no 11.794/2008 – lei que estabelece procedimentos para o uso científico de animais.

5.2.2 Metodologias para identificação e correção da multicolinearidade

As covariáveis dos modelos foram agrupadas em uma matriz **C**, as análises de correção de fenótipo, extração de fatores e de componentes principais foram realizadas nesta matriz, logo após inseridas no banco de dados para as análises de estimação de componentes e predição dos valores genéticos, bem como demais análises exploratórias. Para quantificar e determinar a força das colinearidades, a metodologia de identificação dos índices de condição foi utilizada, em que $10 \leq IC \leq 30$, foi considerada como uma relação de quase dependência, $IC > 30$ como indicativos de colinearidades fortes de acordo com a metodologia proposta por Dias (2008). Para os cálculos de determinação dos autovalores e autovetores da matriz de covariáveis foi utilizado o pacote OLSRR (HEBBALI, 2018) do software R (R TEAM CORE, 2019).

A outra metodologia para identificação da multicolinearidade foi a análise dos fatores de inflação de variância (FIV), este método de identificação da multicolinearidade mede o quanto as variações dos coeficientes de regressão estimados estão infladas se comparados com as variáveis preditoras que não

são linearmente relacionados. Calculado por: $FIV = \frac{1}{1-R_i^2}$ em que, R_i^2 é o coeficiente de correlação múltipla da regressão linear das variáveis explanatórias (X). Quando as variáveis de X não são colineares temos um $R_i^2 = 0$ e consequentemente o $FIV=1$, quando $R_i^2 \neq 0$ o $FIV>1$. Para este estudo foram considerados como indicação de colinearidade $FIV > 10$. Os FIV foram calculados antes e após a aplicação de correção das metodologias dos quadrados mínimos e da regressão de cumeieira.

5.2.2.1 Quadrados mínimos ordinários

O método dos quadrados mínimos ordinários também é chamado de Ordinary Least Square (OLS) quando se considera que os resíduos não são correlacionados e há homocedasticidade. Tendo em vista o modelo linear: $y = X\beta + e$, em que, e é o vetor de observações, X é a matriz de incidência dos efeitos fixos, β é o vetor dos efeitos fixos e e é o vetor de erros aleatórios. O estimador dos quadrados mínimos ordinários (OLS) é obtido por: $\hat{\beta} = (X'X)^{-1}X'y$. O fenótipo original foi então desviado do ajuste pelos quadrados mínimos e inserido no banco de dados para o processamento da estimação dos componentes de variância e da avaliação genética.

As covariáveis dos modelos foram agrupadas em uma matriz C , assim, apesar do estimador de mínimos quadrados de $\hat{\beta}$ ser o estimador linear não viesado de mínima variância, seu quadrado médio residual ainda será grande se existirem colinearidades entre as variáveis explanatórias (DIAS,2008).

5.2.2.2 Regressão de Cumeieira

Considerando que o estimador dos quadrados mínimos ordinários é obtido por $\hat{\beta} = (X'X)^{-1}X'y$, as estimativas são não-viesadas se $E(\hat{\beta}) = \beta$. A matriz de covariâncias é dada por: $V(\hat{\beta}) = \sigma^2 R^{-1}$. Assumindo que y foi padronizado ($\sigma^2 = 1$), tem-se que $V(\hat{b}_i) = \frac{1}{1-R_i^2}$. Em que $\hat{b}_i$ é o i -ésimo coeficiente de regressão contido em β ; R_i^2 é o coeficiente de determinação, obtido através da regressão de y_i em x_i neste caso, essa variância é o fator de

inflação de variância (FIV). É sabido que R_i^2 tende a ser próximo de 1 então o FIV tende a ser grande. Por exemplo, supondo que o $FIV=10$, a solução para o R_i^2 é 0,90, então, em qualquer situação que o $R_i^2 \geq 0,90$ entre uma variável independente e as demais há multicolinearidade.

A diagonal da matriz $X'X$ recebe um pequeno valor de k : $\beta^o = (X'X + kI)^{-1}X'Y$. Em que k corresponde ao parâmetro de cumeeira “ridge” sendo $k > 0$, e I é a matriz identidade associada. O método de determinação de k foi adaptado de Roso et al.(2005a), Dias et al. (2011) e Bertoli et al. (2015) e, estimado como $k_i = \theta [(FIV_i / FIV_{max})]$, em que i é o i -ésimo elemento da matriz diagonal K , FIV_i é o i -ésimo fator de inflação de variância (FIV) da i -ésima variável preditora e FIV_{max} é o valor máximo de todos os FIV_i . Para a escolha do valor de θ , foi realizado o procedimento de *bootstrap*, originalmente proposto por Efron (1979) e descrito por Roso et al. (2005). Este é um processo iterativo, com o valor de θ variando de 0 a 1 com incrementos de 0,00001 em metodologia adaptada de Dias et al. (2011). O valor final de k (parâmetro de cumeeira k) utilizado neste estudo levou em consideração então dois fatores, os valores foram plotados de forma gráfica em que o FIV_{max} fosse abaixo de 10 e então verificado o menor valor de viés associado ao parâmetro de cumeeira. O viés da estimativa foi dado por, $E(\beta^o - \beta) = [(X'X + kI)^{-1}X'X - I]\beta$ e a matriz de (co)variância: $V(\beta^o) = (X'X + kI)^{-1}X'X(X'X + kI)^{-1}$.

5.2.2.3 Análise de Fatores

O principal objetivo da AF é encontrar um conjunto de fatores que formem uma combinação linear das variáveis originais ou da matriz de correlações, se houver a presença de alta correlação entre as variáveis as mesmas serão combinadas para formar um fator, esse processo se repete para todas as variáveis incluídas na análise. No estudo em questão a AF foi aplicada na matriz das covariáveis (**C**), para melhorar a interpretação da solução inicial da análise fatorial e retirar a possível correlação entre os fatores foi aplicada a rotação varimax. O conjunto de fatores que explicaram mais de 90% da variância foram substituídos no modelo de análise dos componentes de (co)variância, para a estimação dos principais parâmetros genéticos. Para a extração dos fatores foi

utilizada a função Factanal do pacote STATS do software estatístico R (R CORE TEAM, 2019).

5.2.2.4 Análise de Componentes Principais

Na regressão de componentes principais (PCA), transformou-se as variáveis independentes em seus componentes principais: $\mathbf{C}'\mathbf{C} = \mathbf{P}\mathbf{D}\mathbf{P}' = \mathbf{Z}'\mathbf{Z}$. Em que $\mathbf{D}$ representa uma matriz diagonal dos autovalores de $\mathbf{C}'\mathbf{C}$, $\mathbf{P}$ é o auto vetor da matriz $\mathbf{C}'\mathbf{C}$ e $\mathbf{Z}$ é a matriz dos dados (com estrutura similar a $\mathbf{C}$) constituída pelos principais componentes. $\mathbf{P}$ é ortogonal de modo que $\mathbf{P}'\mathbf{P} = \mathbf{I}$. Novas variáveis $\mathbf{Z}$ foram criadas, como médias ponderadas das variáveis originais. Como essas novas variáveis são componentes principais, suas correlações entre si são todas zero. Para tanto, foi medido a adequabilidade dos dados, pela utilização do critério de Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett (FERREIRA, 2011). Assim como na análise de fatores os componentes que explicaram mais de 90% da variância foram adicionados aos modelos de análise em substituição as covariáveis da matriz $\mathbf{C}$ e as análises foram processadas.

5.2.3 Modelos para estimação dos componentes de (co)variância

Os modelos foram divididos para as características pré e pós desmama. Para o peso a desmama:

$$y_{ijklm} = GC_i + CIMP_j + \beta_1 IDA + a_k + dam_l + pe_m + e_{ijklm}$$

$$y_{ijklm} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 A + \beta_3 B + \beta_4 C + \beta_5 Am + \beta_6 Bm + \beta_7 Cm + \beta_8 NA \\ + \beta_9 NB + \beta_{10} NC + \beta_{11} AB + \beta_{12} AC + \beta_{13} BC + \beta_{14} Htm + a_k + dam_l \\ + pe_m + e_{ijklm}$$

$$y_{ijklm}^* = GC_i + CIMP_j + \beta_1 IDA + a_k + dam_l + pe_m + e_{ijklm}$$

$$y_{ijklm}^* = GC_i + CIMP_j + \beta_0 INT + \beta_1 IDA + a_k + dam_l + pe_m + e_{ijklm}$$

$$y_{ijklm} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 F1 + \beta_3 F2 + \beta_4 F3 + \beta_5 F4 + \beta_5 F5 + a_k + dam_l + pe_m \\ + e_{ijklm}$$

$$y_{ijklm} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 F1 + \beta_3 F2 + \beta_4 F3 + \beta_5 F4 + \beta_5 F5 + a_k + dam_l + pe_m + e_{ijklm}$$

$$y_{ijklm} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 PC1 + \beta_3 PC2 + \beta_4 PC3 + \beta_5 PC4 + \beta_5 PC5 + a_k + dam_l + pe_m + e_{ijklm}$$

Para as características pós desmama (GP e P14):

$$y_{ijkl} = GC_i + CIMP_j + \beta_1 IDA + a_k + e_{ijkl}$$

$$y_{ijkl} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 A + \beta_3 B + \beta_4 C + \beta_5 Am + \beta_6 Bm + \beta_7 Cm + \beta_8 NA + \beta_9 NB + \beta_{10} NC + \beta_{11} AB + \beta_{12} AC + \beta_{13} BC + \beta_{14} Htm + a_k + e_{ijkl}$$

$$y_{ijkl}^* = GC_i + CIMP_j + \beta_1 IDA + a_k + e_{ijkl}$$

$$y_{ijkl}^* = GC_i + CIMP_j + \beta_1 IDA + a_k + e_{ijkl}$$

$$y_{ijkl} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 F1 + \beta_3 F2 + \beta_4 F3 + \beta_5 F4 + a_k + e_{ijkl}$$

$$y_{ijkl} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 F1 + \beta_3 F2 + \beta_4 F3 + \beta_5 F4 + a_k + e_{ijkl}$$

$$y_{ijkl} = GC_i + CIMP_j + \beta_1 IDA + \beta_2 PC1 + \beta_3 PC2 + \beta_4 PC3 + \beta_5 PC4 + \beta_5 PC5 + a_k + e_{ijkl}$$

Em que, y_{ijklm} e y_{ijkl} representam os vetores de informações fenotípicas; y_{ijklm}^* e y_{ijkl}^* representam os vetores de informações fenotípicas ajustas para os respectivos métodos de correção; GC_i representa o efeito do i-ésimo grupo de contemporâneo; $CIMP_j$ representa o efeito da j-ésima classe de idade da mãe ao parto; β 1 a 14 são os coeficientes de regressão para as covariáveis: IDA (idade à mensuração da característica); INT intercepto da regressão de cumeeira; A, B e C são os efeitos genéticos aditivos diretos associados às composições raciais individuais; Am, Bm e Cm são os efeitos genéticos aditivos diretos associados às composições raciais maternas; NA, NB, NC, AB, AC e BC são os efeitos das heterozigoses diretas; Htm representa a heterozigoses materna total; a_k é o efeito aleatório de animal e e_{ijklm} e e_{ijkl} representam os termos aleatórios residuais.

Os grupos contemporâneos de desmama (GCDES) foram obtidos pela concatenação de fazenda + ano de nascimento (ou safra) + sexo + grupo de manejo à desmama (GMANDES). Os grupos de contemporâneos para o ganho de peso (GCGP) e os de sobreano (GC14) foram obtidos pela concatenação de fazenda + ano de nascimento (ou safra) + fazenda de sobreano + grupo de manejo ao sobreano (GMAN14). Ressalta-se que o grupo de manejo à desmama (GMANDES) não integrou a concatenação para formação dos grupos de

contemporâneos de sobreano (GC14 e GCGP), ou seja, para as características de sobreano, o GMANDES foi incluído, no modelo, como efeito aleatório não correlacionado. As razões para utilização do GMANDES como efeito aleatório não correlacionado são bem descritas em Santana et al. (2013) e Pedrosa et al. (2014).

Assumindo a notação matricial o modelo para PD pode ser representado como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_d + \mathbf{Z}_2\mathbf{a}_m + \mathbf{M}\mathbf{pe} + \mathbf{e}$$

Em que, $\mathbf{y}$ é o vetor de observações fenotípicas; $\boldsymbol{\beta}$ é o vetor de solução dos efeitos fixos; $\mathbf{X}$ é a matriz de incidência dos efeitos fixos; $\mathbf{a}_d$ e $\mathbf{a}_m$ são, respectivamente, o vetor de solução para o efeito genético aditivo direto e o vetor de solução para o efeito genético aditivo materno; $\mathbf{pe}$ é o vetor de solução para o efeito aleatório de ambiente permanente materno; $\mathbf{Z}_1$, $\mathbf{Z}_2$ e $\mathbf{M}$ são matrizes de incidência para os efeitos aleatórios genéticos aditivos, direto e materno, bem como o ambiente permanente materno, respectivamente; $\mathbf{e}$ é o vetor de termos aleatórios residuais associado às observações fenotípicas.

Para as características GP e P14:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_d + \mathbf{W}\mathbf{g} + \mathbf{e}$$

Em que, $\mathbf{y}$ é o vetor de observações fenotípicas; $\boldsymbol{\beta}$ é o vetor de solução dos efeitos fixos; $\mathbf{X}$ é a matriz de incidência dos efeitos fixos; $\mathbf{a}_d$ é o vetor de solução para o efeito genético aditivo direto; $\mathbf{g}$ é o vetor de solução para o efeito aleatório não correlacionado do grupo de manejo a desmama; $\mathbf{Z}_1$ e $\mathbf{W}$ são matrizes de incidência para os efeitos aleatórios genéticos aditivos diretos; $\mathbf{e}$ é o vetor de termos aleatórios residuais associado às observações fenotípicas.

Para ambos os modelos foi assumido que $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$; a variância de $\mathbf{y}$ é dada pela matriz $\mathbf{R}$ que assumiu os componentes de (co)variância para cada modelo especificamente (definida abaixo). Assumiu-se, ainda, que $\mathbf{a}_d|\mathbf{A}, \sigma_{a_d}^2 \sim N(0, \mathbf{A}\sigma_{a_d}^2)$; $\mathbf{a}_m|\mathbf{A}, \sigma_{a_m}^2 \sim N(0, \mathbf{A}\sigma_{a_m}^2)$; $\mathbf{pe}|\mathbf{I}, \sigma_{pe}^2 \sim N(0, \mathbf{I}\sigma_{pe}^2)$; $\mathbf{g}|\mathbf{I}, \sigma_g^2 \sim N(0, \mathbf{I}\sigma_g^2)$ e $\mathbf{e}|\mathbf{I}, \sigma_e^2 \sim N(0, \mathbf{I}\sigma_e^2)$, desta forma $\mathbf{y}|\boldsymbol{\beta}, \mathbf{a}_d, \mathbf{a}_m, \mathbf{pe}, \mathbf{g}, \sigma_{a_d}^2, \sigma_{a_m}^2, \sigma_{a_d a_m}, \sigma_{pe}^2, \sigma_g^2, \sigma_e^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{R})$. A matriz $\mathbf{R}$ foi pressuposta por:

$$R = V \begin{bmatrix} a_d \\ a_m \\ pe \\ g \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_{a_d}^2 & A\sigma_{a_d a_m} & 0 & 0 & 0 \\ A\sigma_{a_m a_d} & A\sigma_{a_m}^2 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{pe}^2 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_g^2 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

Os componentes de (co)variâncias foram estimados pelo método da Máxima Verossimilhança Restrita (REML), utilizando-se o programa REMLF90 (Misztal, 2002). Para a predição dos valores genéticos dos animais, utilizou-se o programa BLUPF90 (MISZTAL et al., 2002).

5.2.4 Comparações dos modelos

Para cada metodologia de análise foi avaliado o valor genético dos animais e posteriormente foram comparados através dos indicadores de qualidade a saber: quadrado médio do erro de predição (QMEP), foi calculado por meio da equação: $QMEP = \frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2$; Porcentagem quadrática de viés (PQV), foi dada por, $PQV = (\sum_{t=1}^n (\hat{y}_t - y_t)^2 / \sum_{t=1}^n (\hat{y}_t)^2) \times 100$; Desvio médio absoluto (DMA), calculado por: $DMA = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|$; Correlação (COR) $r = COV(\hat{y}_t - y_t) / \sqrt{V(\hat{y}_t) \times V(y_t)}$ e correlação do erro (CORE) $r = COV(y_t - \varepsilon) / \sqrt{V(y_t) \times V(\varepsilon)}$ em que, n é o número de observações, y_t corresponde a informação fenotípica observada, $\hat{y}_t$ é a informação fenotípica predita e ε é o erro de predição associado aos modelos.

A estimativa do coeficiente de correlação de Pearson entre os valores genéticos obtidos pelas diferentes metodologias aplicadas foi calculado considerando oito cenários : 1- valores genéticos obtidos por meio da análise SC, 2 - valores genéticos obtidos por meio da análise com covariáveis REML, 3- valores genéticos obtidos por meio da análise QM, com a predição baseada no fenótipo original e os componentes de (co)variância estimados pelo mesmo método corrigido que foi denominado QMP, 4- valores genéticos obtidos por meio da análise QM, com a predição baseada no fenótipo ajustado pela metodologia QM e os componentes de (co)variância estimados pelo mesmo método que foi denominada QMA, valores genéticos obtidos por meio da análise RC, com a predição baseada no fenótipo original e os componentes de

(co)variância estimados pelo mesmo método corrigido que foi denominada RCP, 6 - valores genéticos obtidos por meio da análise RC, com a predição baseada no fenótipo ajustado pela metodologia RC e os componentes de (co)variância estimados pelo mesmo método que foi denominada RCA, 7- valores genéticos obtidos por meio da análise AF e 8- valores genéticos obtidos por meio da análise PCA.

5.3 Resultados

Para o PD (Tabela 2) os resultados referentes aos autovalores 0,0085; 0,0126; 0,0144; 0,0294; 0,0574; apresentaram os respectivos IC associados: 28,4453; 23,3867; 21,8272; 15,2934; 10,9464, indicando relações de quase dependências, ou seja, estes resultados apontam um indicativo de colinearidade fraca. Dias et al. (2012), em estudos com animais Montana, encontraram dois autovalores com dependências fortes para a mesma característica em questão.

Para a característica GP (Tabela 2) os autovalores 0,0011; 0,0066; 0,0106; 0,0294 e 0,0503; apresentaram os respectivos IC associados: 79,4855; 32,9161; 26,0289; 15,6217 e 11,9603, os dois primeiros autovalores apresentaram um IC acima de 30 caracterizando um indicativo forte de colinearidade, já os três últimos indicaram apenas relações de quase dependências. Yakubu (2010), encontraram índices de condição que indicaram fortes dependências para variáveis de crescimento corporal, em uma população africana de bovinos *White Fulani*, os autores evidenciaram a multicolinearidade através desse método. Padrão semelhante foi observado neste estudo para a característica P14 (Tabela 2), em que, os autovalores 0,0013; 0,0067; 0,0109; 0,0295 e 0,0509; apresentaram os respectivos IC associados: 75,3025; 32,8515; 25,7231; 15,5936 e 11,8795.

Tabela 2 – Autovalores, Índice de Condição e FIV

COV	PD				GP				P14			
	EV	IC	FIVQM	FIVRC	EV	IC	FIVQM	FIVRC	EV	IC	FIVQM	FIVRC
A	2,245	1,750	11,071	0,545	2,310	1,764	19,307	0,311	2,306	1,765	18,224	0,337
B	1,523	2,125	9,046	0,708	1,395	2,271	14,605	0,491	1,395	2,269	13,989	0,518
C	1,134	2,463	40,678	0,067	1,033	2,639	52,871	0,048	1,036	2,633	52,960	0,047
Am	0,906	2,755	7,862	0,674	0,892	2,840	14,409	0,485	0,891	2,840	13,442	0,516
Bm	0,749	3,030	13,661	0,415	0,677	3,260	22,157	0,333	0,676	3,260	21,123	0,348
Cm	0,243	5,319	11,937	0,480	0,234	5,549	19,124	0,363	0,238	5,496	18,364	0,381
NA	0,106	8,056	5,587	1,095	0,091	8,872	15,402	0,549	0,092	8,843	14,078	0,598
NB	0,094	8,552	10,135	0,653	0,080	9,453	23,841	0,315	0,084	9,223	21,666	0,354
NC	0,057	10,946	22,873	0,281	0,050	11,960	19,643	0,419	0,051	11,880	19,343	0,422
AB	0,029	15,293	5,236	1,213	0,029	15,652	5,055	1,324	0,030	15,594	4,924	1,354
AC	0,014	21,827	16,825	0,359	0,011	26,029	28,220	0,233	0,011	25,723	28,373	0,232
BC	0,013	23,387	5,673	0,877	0,007	32,916	7,873	0,862	0,007	32,852	8,102	0,858
HTm	0,009	28,445	2,326	1,444	0,001	79,486	1,701	1,241	0,001	75,303	1,699	1,249

COV= covariável, PD= peso à desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou sobreano, EV= auto valor, IC= índice de condição, FIVQM= fator de inflação de variância pelo método dos quadrados mínimos, FIVRC= fator de inflação de variância pelo método da regressão de cumeieira, A= Composição racial do animal do tipo biológico adaptado, B= Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

Os fatores de inflação de variância (FIV) para os métodos de RC apresentaram uma diminuição significativa evidenciando a correção da multicolinearidade, Dias et al. (2008) reportou padrão semelhante para as características de crescimento em bovinos Montana. Nas Tabelas 3,4 e 5 se encontram os coeficientes de regressão para as covariáveis da matriz, na regressão de cumeeira algumas iterações foram necessárias para um ajuste na estabilidade do β da regressão e consequentemente dos valores de k. Para o PD a estabilização destes coeficientes se deu por volta da 7 iteração, a diferença média absoluta entre os valores da 7ª e da 6ª iteração foram muito próximos de zero em alguns casos (0.0001), mas na grande maioria foram nulas as diferenças, em outras palavras, o erro máximo entre as soluções para essa característica foi equivalente a 10^{-5}

Tabela 3 – Coeficientes de regressão e valores dos elementos k da matriz diagonal, para a característica PD.

COV	β_{QM}	β_{RC1}	β_{RC2}	β_{RC3}	β_{RC4}	β_{RC5}	β_{RC6}	β_{RC7}	$\beta_{RC7}-\beta_{RC6}$	k
A	0,525	0,460	-0,003	0,001	0,001	0,002	0,002	0,002	0,000	0,007
B	0,233	0,178	-0,018	-0,012	-0,010	-0,002	-0,010	-0,010	0,000	0,006
C	-0,431	-0,111	0,000	-0,005	-0,006	0,001	-0,006	-0,007	0,001	0,027
Am	-0,472	-0,423	0,008	0,007	0,007	0,003	0,007	0,007	0,000	0,005
Bm	-0,243	-0,200	0,013	0,009	0,008	0,001	0,008	0,008	0,000	0,009
Cm	-0,182	-0,159	0,006	0,003	0,003	0,000	0,003	0,003	0,000	0,008
NA	0,001	0,034	0,001	-0,001	-0,001	-0,003	-0,001	-0,001	0,000	0,004
NB	0,074	0,091	0,012	0,010	0,010	0,002	0,010	0,010	0,000	0,007
NC	0,513	0,328	0,005	0,009	0,010	0,002	0,010	0,011	0,001	0,015
AB	0,006	0,019	0,009	0,011	0,012	0,008	0,012	0,012	0,000	0,004
AC	0,187	0,027	0,006	0,009	0,009	0,002	0,009	0,009	0,000	0,011
BC	0,268	0,087	0,002	0,008	0,010	0,003	0,011	0,011	0,001	0,004
HTm	0,188	0,178	-0,002	-0,002	-0,001	-0,001	-0,001	-0,001	0,000	0,002

COV= covariável, PD= peso à desmama, β_{QM} = coeficiente de regressão pelo método dos quadrados mínimos, β_{RC1} = coeficiente de regressão pelo método da regressão de cumeeira, A=Composição racial do animal do tipo biológico adaptado, B= Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

Para o GP (Tabela 4) a estabilização destes coeficientes se deu por volta da 4ª iteração, a diferença média absoluta entre os valores da 3ª e da 4ª iteração foram nulos. O mesmo aconteceu para o P14, porém foi necessária uma iteração a mais para que a diferença fosse zerada. Alguns autores associaram um número fixo de iterações (DIAS et al., 2012; PETRINI et al., 2012) e reportaram valores semelhantes na estabilização dos β para a regressão de cumeieira.

Tabela 4 – Coeficientes de regressão e valores dos elementos k da matriz diagonal, para a característica GP.

COV	β_{QM}	β_{RC1}	β_{RC2}	β_{RC3}	β_{RC4}	$\beta_{RC3} - \beta_{RC4}$	k
A	0,132	0,065	0,020	0,022	0,022	0,000	0,010
B	0,129	0,063	0,001	0,001	0,001	0,000	0,008
C	0,659	0,262	0,008	0,008	0,008	0,000	0,027
Am	0,044	0,055	-0,011	-0,011	-0,011	0,000	0,007
Bm	0,025	0,034	0,003	0,003	0,003	0,000	0,012
Cm	-0,264	-0,190	-0,009	-0,009	-0,009	0,000	0,010
NA	0,048	0,021	-0,012	-0,012	-0,012	0,000	0,008
NB	0,053	0,023	0,002	0,002	0,002	0,000	0,012
NC	-0,211	-0,067	-0,001	0,000	0,000	0,000	0,010
AB	0,079	0,081	-0,010	-0,011	-0,011	0,000	0,003
AC	0,118	0,218	0,002	0,002	0,002	0,000	0,015
BC	0,254	0,386	0,009	0,009	0,009	0,000	0,004
HTm	-0,038	-0,040	0,006	0,006	0,006	0,000	0,001

COV= covariável, GP= ganho de peso da desmama ao sobreano, β_{QM} = coeficiente de regressão pelo método dos quadrados mínimos, β_{RC1} = coeficiente de regressão pelo método da regressão de cumeieira e suas respectivas iterações, A=Composição racial do animal do tipo biológico adaptado, B= Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

Tabela 5 – Coeficientes de regressão e valores dos elementos k da matriz diagonal, para a característica P14.

COV	β_{QM}	β_{RC}	β_{RC2}	β_{RC3}	β_{RC4}	β_{RC5}	$\frac{\beta_{RC4} - \beta_{RC5}}{\beta_{RC5}}$	k
A	0,401	0,278	0,010	0,012	0,013	0,013	0,000	0,010
B	0,454	0,313	-0,003	-0,004	-0,004	-0,004	0,000	0,008
C	0,620	0,250	0,001	0,004	0,005	0,005	0,000	0,029
Am	-0,225	-0,151	-0,002	-0,003	-0,003	-0,003	0,000	0,007
Bm	-0,166	-0,080	-0,003	-0,005	-0,005	-0,005	0,000	0,012
Cm	-0,425	-0,271	0,005	0,006	0,007	0,007	0,000	0,010
NA	-0,012	0,007	0,004	0,003	0,003	0,002	0,000	0,008
NB	-0,014	0,002	0,009	0,012	0,013	0,013	0,000	0,012
NC	-0,106	0,023	0,008	0,007	0,007	0,007	0,000	0,011
AB	0,110	0,114	-0,005	-0,005	-0,005	-0,005	0,000	0,003
AC	0,238	0,283	-0,006	-0,008	-0,008	-0,008	0,000	0,016
BC	0,255	0,339	0,001	-0,002	-0,003	-0,003	0,000	0,004
HTm	0,133	0,126	0,005	0,007	0,007	0,007	0,000	0,001

COV= covariável, P14= peso aos 14 meses ou ao sobreano, β_{QM} = coeficiente de regressão pelo método dos quadrados mínimos, β_{RC1} = coeficiente de regressão pelo método da regressão de cumeieira e suas respectivas iterações, A=Composição racial do animal do tipo biológico adaptado, B= Composição racial do animal do tipo biológico britânico, C= Composição racial do animal do tipo biológico continental, Am= Composição racial materna do tipo biológico adaptado, Bm= Composição racial materna do tipo biológico britânico, Cm= Composição racial materna do tipo biológico continental, NA= heterozigose entre as raças do grupo N e A, NB= heterozigose entre as raças do grupo N e B, NC= heterozigose entre as raças do grupo N e C, AB= heterozigose entre as raças do grupo A e B, AC= heterozigose entre as raças do grupo A e C, BC= heterozigose entre as raças do grupo B e C, HTm= heterozigose materna total.

Fonte: Própria Autoria

No processo iterativo de detecção do melhor valor de k, sob a condicional de valores para FIV <10, o ponto mais marcante de queda (Figura 1) foi selecionado e posteriormente associado ao respectivo k, a ser incluído nas análises. Para o PD o k escolhido estava associado a um valor máximo de FIV de 8,699 e um valor de viés em torno de 7.02%, valor inferior ao relatado (9.77%) por Dias et al. (2008) em animais Montana, Carvalheiro et al. (2006) em uma população de animais cruzados Nelore x Hereford, associou a porcentagem de viés entre 11,1%, para características de crescimento e de carcaças. Para o GP o FIV máximo associado ao k escolhido foi de 9,89 com uma porcentagem de viés de 9,99%. Para o P14 o FIV foi de 9,89 referente ao k escolhido, com uma porcentagem de viés na ordem de 9,74%.

Figura1 – Fator de inflação de variância (FIV) e porcentagem de viés

Fonte: Própria Autoria

Os componentes de (co)variância (Tabela 6) foram similares para todos os métodos, corroborando com os resultados encontrados por Dias et al. (2011), os autores associaram essa similaridade nos valores, devido ao vetor β não interferir na variabilidade dos efeitos aleatórios. Para a característica PD em que as relações de colinearidade detectadas foram fracas, percebe-se que a alteração dos componentes de (co)variância são bem menores, a herdabilidade variou de 0,22 para o modelo em que eram inclusos as covariáveis, até 0,29 na análise de juste pelos quadrados mínimos. Alguns autores ainda associam também que o principal interesse da utilização da regressão de cumeieira é na análise dos coeficientes de regressão individuais e que a análise de predição há um confundimento nas demais metodologias (BERGMANN & HOHENBOKEN, 1995; FREUND & WILSON, 1998).

Tabela 6 – Estimativas dos componentes de (co)variâncias.

Carac	Param	Método					
		SC	REML	QM	RC	AF	PCA
PD	σa^2	190,90	145,80	197,80	183,60	191,80	181,10
	σm^2	183,70	156,20	174,00	170,40	165,10	164,20
	σe^2	391,00	414,00	389,8	396,00	392,40	397,10
	pe	0,55	0,60	0,59	0,59	0,59	0,59
	c^2	0,00	0,00	0,00	0,00	0,00	0,00
	h^2	0,28	0,22	0,29	0,27	0,28	0,27
	covmat	-80,52	-48,66	-71,41	-65,08	-68,08	-65,39
	h^2_{mat}	0,27	0,23	0,25	0,25	0,24	0,24
	h^2_{tot}	0,24	0,23	0,26	0,25	0,25	0,24
GP	σa^2	110,50	79,39	129,30	125,90	85,67	84,35
	σe^2	488,90	507,60	479,90	482,00	503,70	504,70
	σgm^2	113,80	104,70	117,60	117,80	104,90	105,10
	d^2	0,16	0,15	0,16	0,16	0,15	0,15
	h^2	0,15	0,11	0,18	0,17	0,12	0,12
P14	σa^2	299,70	252,90	357,20	346,40	261,60	258,70
	σe^2	535,00	563,60	509,80	512,70	558,90	557,60
	σgm^2	74,37	76,55	83,42	80,25	76,56	73,83
	d^2	0,08	0,09	0,09	0,09	0,09	0,08
	h^2	0,33	0,28	0,38	0,37	0,29	0,29

Fonte: Própria Autoria

No que se refere a classificação dos modelos (Tabela 7), de acordo com os critérios de comparação, se for considerado apenas as análises SC e REML, percebe-se que o modelo que inclui as covariáveis é o que mais se ajusta aos dados para todas as características, pois os valores encontrados foram inferiores. Ao comparar os demais métodos, tem-se que a análise PCA e a AF ajustaram melhor os dados para todas as características. Schoeman et al. (2002) encontraram valores melhores para a metodologia da RC comparada com a PCA em uma população de bovinos cruzados.

Tabela 7 – Critérios de comparação de modelos.

Carac	Param	Método					
		SC	REML	QM	RC	AF	PCA
PD	AIC	346060,75	345024,77	346797,58	346411,51	345601,11	345328,39
	"-2logL"	346050,75	345014,77	346787,58	346401,51	345591,11	345318,39
GP	AIC	593487,77	593053,36	594000,89	593966,53	593067,90	593103,73
	"-2logL"	593481,77	593047,36	593994,89	593960,53	593061,90	593097,73
P14	AIC	638249,48	637798,6	639392,06	639136,12	637854,93	637758,74
	"-2logL"	638243,48	637792,6	639386,06	639130,12	637848,93	637752,74

COV= covariável, PD= peso á desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou ao sobreano, AIC= critério de Akaike, -2logL= logaritmo da verossimilhança, SC= modelo sem covariáveis, REML = modelo com as covariáveis, QM= quadrados mínimos, RC= regressão de cumeieira, AF = análise de fatores, PCA = análise de componentes principais.

Fonte: Própria Autoria

Na Tabela 8 estão os indicadores de qualidade dos métodos aplicados, o maior QMEP encontrado foi de 987.39 relativo à metodologia QMA consequentemente os valores de PQV, DMA e COR também foram elevados, caracterizando então o pior modelo para a característica PD. Para as demais características os parâmetros demonstraram certa similaridade na magnitude dos seus valores. Para o GP, a metodologia do QMP apresentou o menor QMEP, PQV e DMA, já as estimativas de correlação entre o fenótipo predito e o erro de predição foram as mesmas. Para o P14, em que as relações de

multicolinearidade eram mais severas, a metodologia que se destacou foi o QMP também se mostrou mais eficaz, com o menor valor de QMEP, baixo valor de PQV em conjunto com as metodologias RCA e RCP. Ainda existem poucos estudos que abordam este tema e suas implicações na avaliação genética de animais compostos.

Tabela 8 – Indicadores de qualidade dos métodos.

CARAC.	MÉTODO	QMEP	PQV	DMA	COR ($y, \hat{y}_t$)	CORE (y, ε)
PD	SC	261,10	0,61	12,26	0,91	0,62
	REML	294,21	0,69	13,02	0,90	0,62
	QMA	987,39	2,31	27,21	0,88	0,66
	QMP	260,61	0,61	12,25	0,91	0,62
	RCA	266,80	0,62	12,42	0,91	0,62
	RCP	266,44	0,62	12,39	0,91	0,62
	AF	263,37	0,62	12,32	0,91	0,62
	PCA	270,24	0,63	12,48	0,91	0,62
GP	SC	409,25	4,88	15,07	0,91	0,50
	REML	443,25	5,29	15,68	0,90	0,50
	QMA	391,61	6,34	14,76	0,91	0,50
	QMP	389,10	4,64	14,70	0,91	0,50
	RCA	440,34	5,25	15,63	0,90	0,50
	RCP	439,81	5,24	15,62	0,90	0,50
	AF	436,16	5,20	15,55	0,90	0,50
	PCA	437,77	5,22	15,58	0,90	0,50
P14	SC	376,49	0,46	14,82	0,94	0,49
	REML	419,44	0,51	15,65	0,93	0,48
	QMA	337,02	0,53	14,04	0,95	0,49
	QMP	333,51	0,41	13,95	0,95	0,48
	RCA	339,52	0,41	14,07	0,95	0,48
	RCP	338,61	0,41	14,05	0,95	0,48
	AF	411,94	0,50	15,51	0,93	0,48
	PCA	411,14	0,50	15,49	0,93	0,48

COV= covariável, PD= peso á desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou ao sobreano, QMEP= quadrado médio do erro de predição, PQV= porcentagem quadrática de viés, DMA= desvio médio absoluto, COR= correlação, CORE= correlação do erro, SC= modelo sem covariáveis, REML = modelo com as covariáveis, QM= quadrados mínimos, RC= regressão de cumeieira, AF = análise de fatores, PCA = análise de componentes principais.

Fonte: Própria Autoria

As correlações entre os valores genéticos (Tabela 9), justificam os resultados demonstrados anteriormente.

Tabela 9 – Correlações dos valores genéticos pelos diferentes métodos.

CARAC	MÉTODO	SC	REML	QMA	QMP	RCA	RCP	AF	PCA
PD	SC	1,00	0,92	0,67	1,00	1,00	1,00	0,95	0,98
	REML		1,00	0,84	0,92	0,92	0,92	0,92	0,94
	QMA			1,00	0,68	0,68	0,68	0,81	0,75
	QMP				1,00	1,00	1,00	0,95	0,98
	RCA					1,00	1,00	0,95	0,98
	RCP						1,00	0,95	0,98
	AF							1,00	0,98
	PCA								1,00
GP	SC	1,00	0,89	0,94	1,00	1,00	1,00	0,86	0,88
	REML		1,00	0,84	0,89	0,89	0,89	0,98	0,99
	QMA			1,00	0,94	0,93	0,93	0,81	0,83
	QMP				1,00	0,99	0,99	0,86	0,88
	RCA					1,00	1,00	0,85	0,87
	RCP						1,00	0,85	0,88
	AF							1,00	1,00
	PCA								1,00
P12	SC	1,00	0,94	0,95	1,00	1,00	1,00	0,94	0,94
	REML		1,00	0,90	0,94	0,94	0,94	0,99	0,99
	QMA			1,00	0,95	0,95	0,95	0,88	0,87
	QMP				1,00	1,00	1,00	0,95	0,94
	RCA					1,00	1,00	0,95	0,95
	RCP						1,00	0,95	0,95
	AF							1,00	0,99
	PCA								1,00

PD= peso á desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou ao sobreano, SC= modelo sem covariáveis, REML = modelo com as covariáveis, QMA= fenótipo ajustado pelo método dos quadrados mínimos, QMP= fenótipo original com componentes ajustados pelo método dos quadrados mínimos, RCA= fenótipo ajustado pelo método da regressão de cumeira, RCP= fenótipo original com componentes ajustados pelo método da regressão de cumeira, AF = análise de fatores, PCA = análise de componentes principais.

Fonte: Própria Autoria

Para o PD, foram encontrados os menores valores de correlações, na ordem de 0,67 entre as metodologias QMA e SC, 0,68 para QMA com QMP, RCA, RCP e 0,75 com PCA. Possivelmente esses baixos valores refletem a instabilidade na eficácia dos métodos de correção em situações de relações de

quase dependência entre as variáveis, ou seja, sem situações nas quais a multicolinearidade não é tão prejudicial. Como para esta característica apenas algumas variáveis apresentaram $FIV > 10$ ou $IC > 30$, outra hipótese aventada para a instabilidade nos métodos seria a de que a aplicação da correção para todas as covariáveis (colineares ou não) causaria tal agravante. Foucault (1999) aplicou o método de regressão de cumeieira parcial e encontrou resultados mais satisfatórios, propondo que, neste caso, não ponderar as variáveis não colineares não causa perturbações nos coeficientes de regressão estimados comparados ao método clássico de regressão de cumeieira.

Para o GP e P12, os resultados demonstram que em casos de fortes dependências entre as variáveis explicativas dos modelos, independentemente do método de correção adotado, os componentes aleatórios não são alterados (DÍAZ, 2013). Alguns autores sugerem que o traço de cumeieira “k” seja fixado (HAWKES, 1976; PASHA; SHAH, 2004; BERTOLI; BRACCINI; ROSO, 2016), otimizando a regressão de cumeieira, porém, outro problema surge, no que se refere a escolha do melhor parâmetro para determinada população, tornando o método empírico.

A partir das abordagens estudadas foi calculada a correlação de posto (correlação de Spearman) para verificar as possíveis alterações na classificação dos animais ao longo das safras (Tabela 10). Todas correlações foram positivas entre os métodos de correção. Para a característica PD no método QMA em que o fenótipo foi ajustado para o método dos quadrados mínimos, os menores valores de correlação foram observados (0,66 e 0,67), indicando uma severa mudança no ranqueamento dos animais. Para a característica GP esse impacto no ranqueamento dos animais foi menor, com magnitude das correlações entre os métodos de média a alta ($>0,80$), indicando que não houve alterações severas no ranqueamento dos animais.

Tabela 10 – Correlações de ranking valores genéticos pelos diferentes métodos.

CARAC	MÉTODO	SC	REML	QMA	QMP	RCA	RCP	AF	PCA
PD	SC	1,00	0,92	0,66	1,00	1,00	1,00	0,94	0,97
	REML		1,00	0,81	0,93	0,93	0,93	0,92	0,95
	QMA			1,00	0,67	0,66	0,66	0,82	0,75
	QMP				1,00	1,00	1,00	0,94	0,97
	RCA					1,00	1,00	0,94	0,97
	RCP						1,00	0,94	0,97
	AF							1,00	0,97
	PCA								1,00
GP	SC	1,00	0,89	0,92	1,00	0,99	0,99	0,84	0,87
	REML		1,00	0,84	0,88	0,89	0,89	0,98	0,99
	QMA			1,00	0,92	0,91	0,91	0,79	0,82
	QMP				1,00	0,99	0,99	0,84	0,87
	RCA					1,00	1,00	0,83	0,86
	RCP						1,00	0,83	0,86
	AF							1,00	0,99
	PCA								1,00
P12	SC	1,00	0,94	0,94	1,00	1,00	1,00	0,95	0,94
	REML		1,00	0,91	0,94	0,94	0,94	0,99	0,98
	QMA			1,00	0,95	0,95	0,95	0,89	0,87
	QMP				1,00	1,00	1,00	0,95	0,94
	RCA					1,00	1,00	0,95	0,94
	RCP						1,00	0,95	0,94
	AF							1,00	0,99
	PCA								1,00

PD= peso á desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou ao sobreano, SC= modelo sem covariáveis, REML = modelo com as covariáveis, QMA= fenótipo ajustado pelo método dos quadrados mínimos, QMP= fenótipo original com componentes ajustados pelo método dos quadrados mínimos, RCA= fenótipo ajustado pelo método da regressão de cumeeira, RCP= fenótipo original com componentes ajustados pelo método da regressão de cumeeira, AF = análise de fatores, PCA = análise de componentes principais.

Fonte: Própria Autoria

O mesmo padrão foi observado ao selecionar 1% (Tabela 11) e 30% (Tabela 12) dos melhores animais, porém, para a característica PD a magnitude das correlações paa o método QMA foram inferiores a (0,80), alterando a classificação dos animais.

Tabela 11 – Correlações dos valores genéticos para 1% dos melhores animais pelos diferentes métodos.

CARAC	MÉTODO	SC	REML	QMA	QMP	RCA	RCP	AF	PCA
PD	SC	1,00	0,94	0,69	1,00	1,00	1,00	0,93	0,97
	REML		1,00	0,83	0,94	0,94	0,94	0,94	0,96
	QMA			1,00	0,69	0,69	0,69	0,85	0,78
	QMP				1,00	1,00	1,00	0,94	0,97
	RCA					1,00	1,00	0,94	0,97
	RCP						1,00	0,94	0,97
	AF							1,00	0,97
	PCA								1,00
GP	SC	1,00	0,88	0,92	1,00	0,99	0,99	0,83	0,86
	REML		1,00	0,82	0,88	0,88	0,88	0,98	0,99
	QMA			1,00	0,92	0,90	0,90	0,76	0,80
	QMP				1,00	0,99	0,99	0,83	0,86
	RCA					1,00	1,00	0,82	0,85
	RCP						1,00	0,82	0,85
	AF							1,00	0,99
	PCA								1,00
P12	SC	1,00	0,94	0,94	1,00	1,00	1,00	0,94	0,94
	REML		1,00	0,90	0,94	0,94	0,94	0,99	0,98
	QMA			1,00	0,94	0,94	0,94	0,88	0,86
	QMP				1,00	1,00	1,00	0,94	0,94
	RCA					1,00	1,00	0,94	0,94
	RCP						1,00	0,94	0,94
	AF							1,00	0,99
	PCA								1,00

PD= peso á desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou ao sobreano, SC= modelo sem covariáveis, REML = modelo com as covariáveis, QMA= fenótipo ajustado pelo método dos quadrados mínimos, QMP= fenótipo original com componentes ajustados pelo método dos quadrados mínimos, RCA= fenótipo ajustado pelo método da regressão de cumeira, RCP= fenótipo original com componentes ajustados pelo método da regressão de cumeira, AF = análise de fatores, PCA = análise de componentes principais.

Fonte: Própria Autoria

Tabela 12 – Correlações dos valores genéticos para 30% dos melhores animais pelos diferentes métodos.

CARAC	MÉTODO	SC	REML	QMA	QMP	RCA	RCP	AF	PCA
PD	SC	1,00	0,93	0,66	1,00	1,00	1,00	0,94	0,97
	REML		1,00	0,81	0,93	0,93	0,93	0,93	0,95
	QMA			1,00	0,67	0,67	0,67	0,83	0,75
	QMP				1,00	1,00	1,00	0,94	0,97
	RCA					1,00	1,00	0,94	0,97
	RCP						1,00	0,94	0,97
	AF							1,00	0,97
	PCA								1,00
GP	SC	1,00	0,89	0,92	1,00	0,99	0,99	0,84	0,87
	REML		1,00	0,84	0,89	0,89	0,89	0,98	0,99
	QMA			1,00	0,92	0,91	0,91	0,79	0,82
	QMP				1,00	0,99	0,99	0,84	0,87
	RCA					1,00	1,00	0,83	0,87
	RCP						1,00	0,83	0,87
	AF							1,00	0,99
	PCA								1,00
P12	SC	1,00	0,94	0,94	1,00	1,00	1,00	0,94	0,94
	REML		1,00	0,91	0,94	0,94	0,94	0,99	0,98
	QMA			1,00	0,95	0,95	0,95	0,88	0,87
	QMP				1,00	1,00	1,00	0,95	0,94
	RCA					1,00	1,00	0,95	0,94
	RCP						1,00	0,95	0,94
	AF							1,00	0,99
	PCA								1,00

PD= peso á desmama, GP= ganho de peso da desmama ao sobreano, P14= peso aos 14 meses ou ao sobreano, SC= modelo sem covariáveis, REML = modelo com as covariáveis, QMA= fenótipo ajustado pelo método dos quadrados mínimos, QMP= fenótipo original com componentes ajustados pelo método dos quadrados mínimos, RCA= fenótipo ajustado pelo método da regressão de cumeeira, RCP= fenótipo original com componentes ajustados pelo método da regressão de cumeeira, AF = análise de fatores, PCA = análise de componentes principais.

Fonte: Própria Autoria

Essas divergências para os valores genéticos encontrados na análise que envolveu a correção pelo método dos quadrados mínimos podem estar ligadas a pressuposição do próprio modelo de regressão, em que os estimadores dos β não tem solução única, causando possíveis confundimentos e infinitas possíveis

soluções. Os métodos AF e PCA foram os mais estáveis e demonstraram menores alterações de ranqueamento dos animais.

5.4 Considerações Finais

As correções para a multicolinearidade efetuadas foram eficientes para ajustar o β das covariáveis, para todos os métodos. A AF e PCA apresentaram resultados coerentes com a correção da problemática deste estudo, estas análises tem um potencial elevado para inclusão nas avaliações genéticas de bancos de dados com problema de multicolinearidade, sugere-se a aplicação destas metodologias em outros estudos, principalmente nos que envolvem características categóricas, visto que, nestes casos a correção do fenótipo via análise de QM ou RC não é possível.

A predição dos valores genéticos dos animais para as características deste estudo foi beneficiada pela análise que consistiu na utilização do fenótipo original, com os componentes ajustados para a análise do QM (QMP). A regressão de cumeeira neste estudo pode não ter trazido grandes benefícios, possivelmente pela empiricidade de estimação do parâmetro “k”. Por mais que os métodos apresentaram diferenças nas análises de comparação de modelos, como o ajuste do β das covariáveis não interfere na predição dos valores genéticos, recomenda-se a aplicação da correção para a estimação dos componentes de variância e dos parâmetros genéticos, cabe ao pesquisador então um maior detalhamento do seu banco de dados para a escolha do modelo parcimonioso mais adequado.

Referências

- BELSLEY, D. A.; KUH, E.; WELSCH, R. E. Identifying influential data and sources of collinearity. **Regression Diagnostics**, 1980.
- BERTOLI, C. D.; BRACCINI, J.; ROSO, V. M. Comparing methodologies to estimate fixed genetic effects and to predict genetic values for an Angus x Nellore cattle population. **Journal of Animal Science**, v. 94, n. 2, p. 500–513, 2016.
- BERTOLI, C. D.; NETO, J. B.; MCMANUS, C.; COBUCI, J. A.; CAMPOS, G. S.; PICCOLI, M. L.; ROSO, V. Modelling non-additive genetic effects using ridge regression for an Angus–Nellore crossbred population. **Animal Production Science**, v. 59, n. 5, p. 823, 2019. Disponível em: <<http://www.publish.csiro.au/?paper=AN17439>>.
- DIAS, M.; LIMA, E.; FUKAHORI, F.; SILVA, V.; RÊGO, M. Cinomose canina : revisão de literatura [Distemper canine : literature review]. **Medicina Veterinária**, v. 6, n. 4, p. 32–40, 2012.
- DIAS, R. A. P. **Diagnóstico e redução da influência da multicolinearidade na estimação de efeitos genéticos aditivos e não-aditivos em uma população de bovinos compostos (Bos taurus x Bos indicus)**. 2008. scola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2008.
- DIAS, R. A. P.; PETRINI, J.; FERRAZ, J. B. S.; ELER, J. P.; BUENO, R. S.; DA COSTA, A. L. L.; MOURÃO, G. B. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. **Livestock Science**, v. 142, n. 1–3, p. 188–194, dez. 2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1871141311002836>>.
- DÍAZ, J. R. Inclusão de efeitos genéticos não aditivos na avaliação de características de crescimento e carcaça em bovinos compostos (Bos taurus x Bos indicus). 2013.
- EFRON, B. Bootstrap methods: Another look at the Jackknife. **Annals of Statistics**. 7:1–26. 1979.
- FOUCAULT, T. Order flow composition and trading costs in a dynamic limit order market. **Journal of Financial markets**, v. 2, n. 2, p. 99–134, 1999.
- HAWKES, J. S. Hawkes, James S. Ridge Estimation in Linear Regression. **CLEMSON UNIV SC DEPT OF MATHEMATICAL SCIENCES**., v. 80, 1976.
- HEBBALI, A. **olsrr: Tools for Building OLS Regression Models**, 2018. . Disponível em: <<https://cran.r-project.org/package=olsrr>>.
- KINGHORN, B. P.; VERCOCOE, P. E. The effects of using the wrong genetic model to predict the merit of crossbred genotypes. **Animal Production**, v. 49, p. 209–216, 1989.
- MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D.

BLUPF90 and related programs(BGF90). in **Proceedings of the 7th World Congress on Genetics Applied to Livestock Production.**, p. 21–22, 2002.

PASHA, G. .; SHAH, M. A. Application of ridge regression to multicollinear data. **Journal of research (Science)**, v. 15, n. 1, p. 97–106, 2004.

PEDROSA, V. B.; ELER, J. P.; FERRAZ, J. B. S.; GROENEVELD, E. Use of weaning management group as a random effect for a more robust estimation of genetic parameters for post-weaning traits in Nellore cattle. **Genetics and Molecular Research**, v. 13, n. 3, p. 7013–7021, 2014.

PETRINI, J.; ANTONIO, R.; DIAS, P.; FERNANDA, S.; PERTILE, N.; ELER, J. P.; BENTO, J.; FERRAZ, S.; MOURÃO, G. B. Degree of multicollinearity and variables involved in linear dependence in additive - dominant models. n. 1, p. 1743–1750, 2012.

PIMENTEL, E. da C. G. **ALTERNATIVAS PARA REDUÇÃO DE EFEITOS DE MULTICOLINEARIDADE EM MODELOS DE AVALIAÇÃO DE EFEITOS GENÉTICOS EM BOVINOS DE CORTE**. 2004. Faculdade de Ciências Agrárias e Veterinárias – Unesp, 2004.

R TEAM CORE. **A language and environment for statistical computing.R Foundation for Statistical Computing, Vienna, Austria**.Vienna, Austria, 2019. . Disponível em: <www.R-project.org/>.

ROSO, V. M.; SCHENKEL, F. S.; MILLER, S. P.; SCHAEFFER, L. R. Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation. **Journal of Animal Science**, v. 83, n. 8, p. 1788–1800, 2005a.

ROSO, V. M.; SCHENKEL, F. S.; MILLER, S. P.; WILTON, J. W. Additive, dominance, and epistatic loss effects on preweaning weight gain of crossbred beef cattle from different Bos taurus breeds. **Journal of Animal Science**, v. 83, n. 8, p. 1780–1787, 2005b.

SANTANA, M. L.; ELER, J. P.; FERRAZ, J. B. S. Alternative contemporary group structure to maximize the use of field records: Application to growth traits of composite beef cattle. **Livestock Science**, v. 157, n. 1, p. 20–27, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.livsci.2013.06.034>>.

SCHOEMAN, S. J.; AZIZ, M. A.; JORDAAN, G. F. The influence of multicollinearity on crossbreeding parameter estimates for weaning weight in beef cattle. **South African Journal of Animal Sciences**, v. 32, n. 4, p. 239–246, 2002.

YAKUBU, A. Fixing multicollinearity instability in the prediction of body weight from morphometric traits of White Fulani cows. **Journal of Central European Agriculture**, v. 11, n. 4, p. 487–492, 2010.

```
#####
#
#
#
#
#
#
#
#
#
#
#####
#
#Preparação dos arquivos de dados com correção para multicolinearidade utilizando as técnicas#
#dos Quadrados Mínimos, Regressão de Cumeira, Análise de Fatores e de Componentes Principais#
#####
##### Criar matriz X que recebe as cováriáveis #####
##### 6-8 composição racial do individuo #####
##### 10-12 composição racial materna #####
##### 13-18 heterozigoses diretas e #####
##### 19 heterozigose materna total #####
##### Y vetor de informações fenotípicas (20) #####
#####

setwd("~/banco1/p12")
unip12 <- read.table("unip12r1_.txt", h=T, sep=" ")
#install.packages("stats")
library(stats)
#install.packages("olsrr")
library(olsrr)
#install.packages("Rcmdr")
library(Rcmdr)
#install.packages("plyr")
library(plyr)
#install.packages("Rcpp")
library(Rcpp)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("mctest")
library(mctest)

#descritiva do banco e dos GC's
P12 <- cbind(length(unip12[,49]), mean(unip12[,49]), min(unip12[,49]),
max(unip12[,49]), sd(unip12[,49]),length(unique(unip12[,39])))
P12
gcs<-count(unip12[,38])
sex<- count(unip12[,46])
hist(unip12[,49])
plot(density(unip12[,49]))

#####
##### Critérios de Classificação dos NABC #####
#####

unip12$n <- unip12$/100
unip12$a <- unip12$a/100
unip12$b <- unip12$b/100
unip12$c <- unip12$c/100
unip12$nABC <- NA
str(unip12)

unip12$nABC[unip12$n>0.60 & unip12$n<0.90]<-'12XXX'
unip12$nABC[unip12$a>0.60 & unip12$a<0.90]<-'X12XX'
unip12$nABC[unip12$b>0.60 & unip12$b<0.90]<-'XX12X'
unip12$nABC[unip12$c>0.60 & unip12$c<0.90]<-'XXX12'

unip12$nABC[unip12$n>=0.90]<-'16000'
unip12$nABC[unip12$a>=0.90]<-'01600'
unip12$nABC[unip12$b>=0.90]<-'00160'
unip12$nABC[unip12$c>=0.90]<-'00016'
```

```

unip12$ABC[(unip12$n>=0.40 & unip12$n<=0.60) & (unip12$a>=0.40 & unip12$a<=0.60)]<-'8800'
unip12$ABC[(unip12$n>=0.40 & unip12$n<=0.60) & (unip12$b>=0.40 & unip12$b<=0.60)]<-'8080'
unip12$ABC[(unip12$n>=0.40 & unip12$n<=0.60) & (unip12$c>=0.40 & unip12$c<=0.60)]<-'8008'
unip12$ABC[(unip12$a>=0.40 & unip12$a<=0.60) & (unip12$b>=0.40 & unip12$b<=0.60)]<-'0880'
unip12$ABC[(unip12$a>=0.40 & unip12$a<=0.60) & (unip12$c>=0.40 & unip12$c<=0.60)]<-'0808'

unip12$ABC[(unip12$n>0.1875 & unip12$n<0.3125) & (unip12$a>0.1875 & unip12$a<0.3125) &
(unip12$b>0.1875 & unip12$b<0.3125) & (unip12$c>0.1875 & unip12$c<0.3125)] <- '4444'
unip12$ABC[(unip12$n>0.1875 & unip12$n<0.3125) & (unip12$a>0.1875 & unip12$a<0.3125) &
(unip12$b>0.4375 & unip12$b<0.5625) & (unip12$c<0.0625)] <- '4480'
unip12$ABC[(unip12$n>0.1875 & unip12$n<0.3125) & (unip12$a>0.4375 & unip12$a<0.5625) &
(unip12$b<0.0625) & (unip12$c>0.1875 & unip12$c<0.3125)] <- '4804'
unip12$ABC[(unip12$n>0.1875 & unip12$n<0.3125) & (unip12$a>0.4375 & unip12$a<0.5625) &
(unip12$b>0.1875 & unip12$b<0.3125) & (unip12$c<0.0625)] <- '4840'
unip12$ABC[(unip12$n<=0.375) & (unip12$a>=0.125 & unip12$a<=0.875) & (unip12$b<=0.75) &
unip12$c<=0.75) & ((unip12$n+unip12$a)<=0.25) & ((unip12$b+unip12$c)<=0.75) &
is.na(unip12$ABC))] <- 'XXXX'

unip12$ABC[is.na(unip12$ABC)] <- '0000'

unip12$ABCN <- NA
unip12$ABCN[unip12$ABC=="0000"] <- 0
unip12$ABCN[unip12$ABC=="12XXX"] <- 1
unip12$ABCN[unip12$ABC=="X12XX"] <- 2
unip12$ABCN[unip12$ABC=="XX12X"] <- 3
unip12$ABCN[unip12$ABC=="XXX12"] <- 4
unip12$ABCN[unip12$ABC=="16000"] <- 5
unip12$ABCN[unip12$ABC=="01600"] <- 6
unip12$ABCN[unip12$ABC=="00160"] <- 7
unip12$ABCN[unip12$ABC=="00016"] <- 8
unip12$ABCN[unip12$ABC=="8800"] <- 9
unip12$ABCN[unip12$ABC=="8080"] <- 10
unip12$ABCN[unip12$ABC=="8008"] <- 11
unip12$ABCN[unip12$ABC=="0880"] <- 12
unip12$ABCN[unip12$ABC=="0808"] <- 13
unip12$ABCN[unip12$ABC=="4444"] <- 14
unip12$ABCN[unip12$ABC=="4480"] <- 15
unip12$ABCN[unip12$ABC=="4804"] <- 16
unip12$ABCN[unip12$ABC=="4840"] <- 17
unip12$ABCN[unip12$ABC=="XXXX"] <- 18

hist(unip12$ABCN[unip12$ABCN!=5 & unip12$ABCN!=0])

freqnabc<- data.frame(table(unip12$ABCN))

#####
##### PREPARAÇÃO DAS MATRIZES E VETORES #####
#####

# Matriz das composições e das heterozigoses
X <- cbind(unip12[,21:23],unip12[,25:27],unip12[,28:33],unip12[,35])
colnames(X)<- c("A","B","C","Am","Bm","Cm","NA","NB","NC","AB","AC","BC","HTm")

# Vetor do fenótipo P12[,48] P12[,49] GP[,50]
Y <- unip12[,49]

# Matriz com a média das colunas de X
meanX <- colMeans(X)

#Centralização da matrix X
Xc <- scale(X,meanX,scale = F)

#Soma dos quadrados de cada coluna
cssX <- colSums(Xc^2)

#Padronização de Xc
sX <- Xc %*% diag(1/(sqrt(cssX)))

#Criando Matriz X'X
XX <- t(sX) %*% sX

# Matriz com a média de Y
meanY <- mean(Y)

```

```

#Centralização de Y
Yc <- scale(Y,meanY,scale = F)

#Criando Matriz XY
XY <- t(sX) %*% Yc

### Grafico Correlação###
install.packages("corrplot")
library(corrplot)
col <- colorRampPalette(c("#67001F", "#B2182B", "#D6604D", "#F4A582", "#FDDBC7",
                           "#FFFFFF", "#D1E5F0", "#92C5DE", "#4393C3", "#2166AC",
                           "#053061"))
corX <- cor(X)
tiff("corpl2.tiff", width = 3500 , height = 2500 , units = "px", res =600, compression =
"lzw" )
corrplot(corX, type = "lower", method = "circle",
          col=col(200),tl.srt = 0, tl.col = "black",addgrid.col=NA,
          diag = F, insig="blank", )
dev.off()
write.table(corX,"cor_pl2.txt",col.names= T,row.names=F,quote=F,sep= " ")

#####
##### ÍNDICE DE CONDIÇÃO PARA DETECÇÃO DA MULTICOLINEARIDADE #####
#####

model <- lm(Y~unipl2[,21] + unipl2[,22] + unipl2[,23] +
            unipl2[,25] + unipl2[,26] + unipl2[,27] +
            unipl2[,28] + unipl2[,29] + unipl2[,30] +
            unipl2[,31] + unipl2[,32] + unipl2[,33] +
            unipl2[,35])

IC <- cbind(ols_eigen_cindex(model))
write.table(IC,"IC_P12.txt",col.names= T,row.names=F,quote=F,sep= " ")

#####
##### OBTENDO AS SOLUÇÕES POR QM #####
#####

#Criando os coeficientes de regressão para cada covariável
betaR <- solve(XX) %*% XY
betaR

#Criando os coeficientes de regressão padronizados pela soma de quadrados
betaQM <- diag(1/(sqrt(cssX))) %*% betaR
betaQM

#Criando os fatores de inflação de variância para os QM (MED,SD,MAX)
FIVQM <- diag(solve(XX))
FIVQM

mFIVQM <- FIVQM/max(FIVQM)
mFIVQM

sFIVQM <- sum(FIVQM)/ncol(XX)
sFIVQM

maxFIVQM <- max(FIVQM)
maxFIVQM

#Criando o fenótipo corrigido por QM e inserindo no banco para análise genética
P12_QM <- Y - (as.matrix(X) %*% betaQM)
unipl2$P12_QM <- P12_QM

#####
##### OBTENDO AS SOLUÇÕES POR RC #####
#####

teta <- seq(0,1, by=0.00001)
teta

#Loop para criacao da matriz dos K's

```

```

K <- matrix(NA,ncol(X),length(teta))
i <- 1
j <- 1

for (j in 1:ncol(K)) {
  for (i in 1:nrow(K)) {
    K[i,j] <- teta[j]*mFIVQM[i]
    i <- i+1
    print(c(i,j))
  }
}

write.table(K, "K_P12.txt", col.names = F , row.names= F, quote=F, sep = " ")
K<- read.table("K_P12.txt")

#Loop para calcular XXIk - retorna com a % de vies e o max FIV para cada K

resFIVRC <- matrix(NA,2,length(teta))
j <- 1
for (j in 1 : ncol(resFIVRC)) {
  XXIk <- XX + diag(K[,j])
  iXXIk <- solve(XXIk)
  H = iXXIk %*% XX
  H2 = H ^ 2
  iBIAS <- (1-(sqrt(sum(H2))/sqrt(ncol(XX)))) * 100
  FIVRC <- diag(iXXIk %*% XX %*% iXXIk)
  resFIVRC[1,j] <- max(FIVRC)
  resFIVRC[2,j] <- iBIAS
}

#Transformação do resFIV em dataframe(colunas)
RESULT <- data.frame(t(resFIVRC))
names(RESULT) <- c("mxFIV","BIAS")
RESULT$K <- seq(1:length(teta))

#ordenação através do menor vies
RESULT <- RESULT[order(RESULT$BIAS),]

plot(RESULT$BIAS,RESULT$mxFIV)

RESULT2 <- RESULT[RESULT$mxFIV < 10,]
plot(RESULT2$BIAS,RESULT2$mxFIV)

locator()

#aproximação do gráfico para escolha do melhor ponto de coordenadas de K

#plot(RESULT$BIAS[RESULT$K >= 2000 & RESULT$K <= 3750],
#RESULT$mxFIV[RESULT$K >= 2000 & RESULT$K <= 3750])
#locator()
#x = 6.370405 y = 8.576043
kfinal <- RESULT[RESULT$BIAS <= 11.5 & RESULT$mxFIV <= 8.3,]
#em que, k = 4014 e BIAS= 11.47457 e mFIV = 8.299167
#4 iteração, k= 2915 eBIAS= 9.894799 e mFIV = 9.749716,

#Plotando grafico de coordenadas de K
tiff("kfinalp12_en_s.tiff", width = 3380 , height = 2500 , units = "px", res =600 ,
compression = "lzw")
par(mar=c(5,5,2,2))
plot(RESULT$BIAS,RESULT$mxFIV, xlab = "BIAS(%)", ylab = "VIF(max)", family="serif",
cex.lab=1.5, cex.axis=1.5, cex.names=1.5,
pch=".", axes=F, xlim = range(0:35), ylim = range(0:45))
axis(1,seq(0,35,5))
axis(2,seq(0,45,5),las=2)
segments(9.894799,9.749716,9.894799,0, lty = 2)
segments(0,9.749716,9.894799,9.749716, lty = 2)
dev.off()

#Substituindo pelo melhor K
XXIk <- XX + diag(K[,4014])
iXXIk <- solve(XXIk)
H = iXXIk %*% XX
H2 = H ^ 2
iBIAS <- (1-(sqrt(sum(H2))/sqrt(ncol(XX)))) * 100

```

```

FIVRC <- diag(iXXIk %*% XX %*% iXXIk)
sFIVRC <- (sum(FIVRC))/ncol(XX)
maxFIVRC <- max(FIVRC)
solr <- iXXIk %*% XY
betaRC <- diag(1/(sqrt(cssX))) %*% solr
vk <- diag(K[,4014])

#Criando o intercepto da analise genetica - int

kcond <- Y - (as.matrix(X) %*% betaRC)
int <- mean(kcond)

#Criando o fenótipo corrigido por RC e inserindo no banco para análise genética

P12_RC <- Y - (as.matrix(X) %*% betaRC)
unip12$P12_RC <- P12_RC
unip12$int <- int

#####
#####          ANÁLISE DE FATORES          #####
#####

#análise de componentes principais com a matriz das covariáveis

loadings(pca)
#Foram escolhidos 6 componentes - representam 93,11% da var#
pca.f <- factanal(Xc, factors = 6, rotation="varimax", scores="regression")

ev <- eigen(Xc)

#componentes para serem inseridos na analise genética
scores<- pca.f$scores

#agrupamento das colunas com os componentes
unip12 <- cbind(unip12,scores)

#####
#####          ANÁLISE DE COMPONENTES PRINCIPAIS          #####
#####

cor<- cor(X)
eigen<- eigen(cor)
lambda<- eigen$values
C<- eigen$vectors
perc<-100*lambda/length(lambda)
cumsum(perc)
score <- data.frame(as.matrix(X)%*%C[,1:6])

unip12<- cbind(unip12,score)

#####
#####          BANCO FINAL PARA ANÁLISE          #####
#####
banalP12 <- unip12[c(1,2,3,39,7,45,21:23,25:33,35,56,49,62,63,64,65:70)]
names(banalP12) <- c("ID_PRODUTO","ID_PAI","ID_MAE",
                    "GC_12", "IDADE12", "CIMP",
                    "A","B","C","Am","Bm","Cm","NA","NB","NC","AB","AC","BC","HTm",
                    "GMANDP12","P12_ORI","QM","RC","INT",
                    "PC1","PC2","PC3","PC4","PC5","PC6")
write.table(banalP12, "banalP12.txt", col.names = F , row.names= F, quote=F, sep = " ")

layout_unip12 <- matrix(NA,ncol(unip12),2)
layout_unip12[,1]<-names(unip12)
layout_unip12[,2]<- 1:ncol(unip12)
write.table(layout_unip12,"layout_unip12.txt",col.names= F,row.names=F,quote=F,sep= " ")

```



```

#Layout para o cartão de parâmetros
layout_banalP12 <- matrix(NA,ncol(banalP12),2)
layout_banalP12[,1]<-names(banalP12)
layout_banalP12[,2]<- 1:ncol(banalP12)
write.table(layout_banalP12,"layout_banalP12.txt",col.names= F,row.names=F,quote=F,sep= "
")

head(banalP12)
tail(banalP12)

#####
#####                                ITERAÇÕES                                #####
#####

setwd ('/home/barbara/it/p12')

#Importando banco
banalp12<- read.table("renf90.dat", h=F)
names(banalp12)
c("RC","GC","CIMP","INT","IDA12","ID_R","GMAND","P12_ORI","ID_PRODUTO")
banalp12<- banalp12[order(banalp12$ID_R),]

#Corrigindo Y
Y <- read.table ("yhat_residual_RC4", h=F)
Y <- Y[order(Y$V1),]
Y <- Y[,4]

#Loop para calcular XXIk - retorna com a % de vies e o max FIV para cada K
# Matriz com a média das colunas de X

meanX <- colMeans(X)
Xc <- scale(X,meanX,scale = F)
cssX <- colSums(Xc^2)
sX <- Xc %*% diag(1/(sqrt(cssX)))
XX <- t(sX) %*% sX
meanY <- mean(Y)
Yc <- scale(Y,meanY,scale = F)
XY <- t(sX) %*% Yc

resFIVRC <- matrix(NA,15,length(teta))
j <- 1
for (j in 1 : ncol(resFIVRC)) {
  XXIk <- XX + diag(K[,j])
  iXXIk <- solve(XXIk)
  H = iXXIk %*% XX
  H2 = H ^ 2
  iBIAS <- (1-(sqrt(sum(H2))/sqrt(ncol(XX)))) * 100
  FIVRC <- diag(iXXIk %*% XX %*% iXXIk)
  resFIVRC[1,j] <- max(FIVRC)
  resFIVRC[2,j] <- iBIAS
  solr <- iXXIk %*% XY
  betaRC <- diag(1/(sqrt(cssX))) %*% solr
  resFIVRC[3:15,j] <- betaRC
}

RESULT <- data.frame(t(resFIVRC))
names(RESULT) <- c("mxFIV","BIAS","B1","B2","B3","B4","B5",
                  "B6","B7","B8","B9","B10","B11","B12","B13")
RESULT$K <- seq(1:100001)

RESULT <- RESULT[RESULT$mxFIV < 10,]

plot(RESULT$BIAS,RESULT$mxFIV)
locator()

kfinal <- RESULT[RESULT$BIAS <= 9.9 & RESULT$mxFIV <= 9.75,]
#em que,          k= 3443 e mFIV = 8.99910, BIAS=          10.70252
#2 iteração , k= 3083 e mFIV = 9.498549, BIAS= 10.40452 3083 9.498599 10.16321
#3 iteração , k= 2787 e mFIV = 9.94955, BIAS= 9.682242
#4 iteração, k= 2915 e mFIV = 9.749716, BIAS= 9.894799

```

```

kfinal[row.names(kfinal)=="2915",c("mxFIV", "BIAS")]

Ks <- cbind(K[,4096],K[,3443],K[,3083],K[,2787],K[,2915])
Ks <- round(Ks,5)

#Substituindo pelo melhor K (lembrar de substituir o k)
XXIk <- XX + diag(K[,2915])
iXXIk <- solve(XXIk)
H = iXXIk %*% XX
H2 = H ^ 2
iBIAS <- (1-(sqrt(sum(H2))/sqrt(ncol(XX)))) * 100
FIVRC <- diag(iXXIk %*% XX %*% iXXIk)
sFIVRC <- (sum(FIVRC))/ncol(XX)
maxFIVRC <- max(FIVRC)
solr <- iXXIk %*% XY
betaRC <- diag(1/(sqrt(cssX))) %*% solr
vk <- diag(K[,2915])

#Criando o intercepto da analise genetica - int
kcond <- Y - (as.matrix(X) %*% betaRC)
int <- mean(kcond)

#Criando o fenótipo corrigido por RC e inserindo no banco para análise genética
Y <- banalp12$P12_ORI
banalp12$RC <- Y - (as.matrix(X) %*% betaRC)
banalp12$INT <- int

banalp12<- banalp12[order(banalp12$GC,banalp12$CIMP,banalp12$ID_R),]

write.table(banalp12, "renf90.dat", col.names = F , row.names= F, quote=F, sep = " ")

#####
#                               /)  (\                               #
#      .-.-((,~~.))_.-,      F                               #
#      \.-.  @@  /-.'      I                               #
#      / ,o--o. \      M                               #
#      ( ( .___. )      #Bárbara Abreu Silva - barbara.abreusilva@usp.br#
#      ) \-----' (      #
#      /          \      #
#####

```